

ETHOS

KWARTALNIK
INSTYTUTU
JANA PAWŁA II KUL
LUBLIN

Nr 143
lipiec-wrzesień 2023

OSWOIĆ SZTUCZNĄ INTELIGENCJĘ

W numerze m.in.

Ks. Marcin FERDYNUS

* Czy mamy prawo
oswoić sztuczną inteligencję?

Mariusz WOJEWODA

* Artificial Intelligence
as a Social Utopia

Grażyna OSIKA

* Komunikacja człowieka
ze sztuczną inteligencją

Jakub GOMUŁKA

* Artificial Intelligence
Applied to Philosophy

Leszek KORPOROWICZ

* Redukcyjne pułapki socjologii
„myślących maszyn”

Renata TOMASZEWSKA

* Raj – piekło – triumf



KOLEGIUM REDAKCYJNE KWARTALNIKA „ETHOS”

Michel BASTIT, Sergio BELARDINELLI, Gerald J. BEYER, Rocco BUTTIGLIONE
Christoph BÖHR, Barbara CHYROWICZ SSpS, John CROSBY, Aimable-André DUFATANYE
Krzysztof DYBCIAK, Tomasz GARBOL, Dariusz GAWIN, Piotr GUTOWSKI, Michael JAMES
Wojciech KACZMAREK, Karol KLAUZA, Stanislav KOŠČ, Ewa Agnieszka LEKKA-KOWALIK
Hubert ŁASZKIEWICZ, Leszek MĄDZIK, Balázs M. MEZEI, Maciej NOWAK, ks. Sławomir NOWOSAD
Antonio PARDO CABALLOS, Furio PESCI, Adam POTKAY, Giovanni SALMERI
Stefan SAWICKI, ks. Jan SOCHOŃ, Manfred SPIEKER, Zbigniew STAWROWSKI
Eleonore STUMP, ks. Alfred M. WIERZBICKI, Linda ZAGZEBSKI, Ireneusz ZIEMIŃSKI

ZESPÓŁ REDAKCYJNY KWARTALNIKA „ETHOS”

ks. Marek SŁOMKA – redaktor naczelny
Jarosław MERECKI SDS – zastępca redaktora naczelnego
Natalia GONDEK – sekretarz redakcji
Tomasz GÓRKA – sekretarz redakcji
Dorota CHABRAJSKA, Mirosława CHUDA, Bernadetta KUCZERA-CHACHULSKA, Patrycja MIKULSKA
Ferdie McDERMOTT (redaktor językowy), Wojciech KRUSZEWSKI (redaktor językowy)

REDAKTORZY TOMU

Ewa Agnieszka LEKKA-KOWALIK, Mariusz WOJEWODA

RECENZENCI TOMU

Maurizio BALISTRERI, Caleb BERNACCHIO, Wojciech J. BOBER, Anna BUGAJSKA, Aleksandra DERRA
ks. Michał DROŻDŻ, Teresa GRABIŃSKA, Marek HETMAŃSKI, ks. Grzegorz HOŁUB, Kazimierz KRZYSZTOFEK
Anna LATAWIEC, Krzysztof MICHALSKI, Józef MŁYŃSKI, Tadeusz MICZKA, Danuta OPACKA-WALASEK
Jan PLESZYCZYŃSKI, Halina RAROT, Mariola SUŁKOWSKA-JANOWSKA, Barry SMITH
Jan STASIEŃKO, Joseph WANG-KATHREIN, Krzysztof WIELECKI, Joanna ZACH

Przygotowanie do druku

Dorota CHABRAJSKA, Mirosława CHUDA, Katarzyna GERYN-ADAMOWICZ, Natalia GONDEK
Tomasz GÓRKA, Patrycja MIKULSKA, Mirella NAWRACAŁA-URBAN

Przygotowanie komputerowe

Konrad NOWAKOWSKI

Projekt okładki

Leszek MĄDZIK

Fotografia na okładce

Robert PRANAGAL

ETHOS jest kwartalnikiem filozoficznym o charakterze interdyscyplinarnym
publikowanym w wersji papierowej (jest to wersja pierwotna).
Fragmenty artykułów dostępne są na stronie www.ethos.lublin.pl.



Ministerstwo Kultury
i Dziedzictwa Narodowego

Dofinansowano ze środków
Ministra Kultury i Dziedzictwa Narodowego
pochodzących z Funduszu Promocji Kultury

ISSN 0860-8024
e-ISSN 2720-5355
www.ethos.lublin.pl

ETHOS

Rok 36
2023 nr 3(143)

OSWOIĆ SZTUCZNĄ INTELIGENCJĘ

- * Od Redakcji – Lęk i nadzieja (A.L.K., M.W.) **5**
- * From the Editors – Fear and Hope (A.L.K., M.W.) **13**
- * JAN PAWEŁ II – Szybki rozwój (Fragmenty Listu apostołskiego do odpowiedzialnych za środki społecznego przekazu, Watykan, 24 I 2005) **23**
- * JOHN PAUL II – The Rapid Development (Extracts from the Apostolic Letter to Those Responsible for Communications, Vatican, 24 January 2005) **33**

KU ZROZUMIENIU SZTUCZNEJ INTELIGENCJI

- * Zbigniew ORBIK – W poszukiwaniu filozoficznych źródeł problematyki sztucznej inteligencji **45**
DOI 10.12887/36-2023-3-143-05
- * Ks. Marcin FERDYNUS – Czy mamy prawo ośwoić sztuczną inteligencję? **67**
DOI 10.12887/36-2023-3-143-06
- * Paweł POLAK, Roman KRZANOWSKI – How to Tame Artificial Intelligence? A Symbiotic AI Model for Beneficial AI **92**
DOI 10.12887/36-2023-3-143-07

SZTUCZNA INTELIGENCJA W SPOŁECZEŃSTWIE LUDZKIM

- * Mariusz WOJEWODA – Artificial Intelligence as a Social Utopia **109**
DOI 10.12887/36-2023-3-143-08
- * Grażyna OSIKA – Komunikacja człowieka ze sztuczną inteligencją. Warunki brzegowe **127**
DOI 10.12887/36-2023-3-143-09

- * Łukasz SAROWSKI, Małgorzata GRUCHOŁA – Robotyka społeczna w perspektywie (nie)antropocentryzmu **154**
DOI 10.12887/36-2023-3-143-10
- * Anna MAJ – Ewolucja inteligencji i jej badanie. Oswajanie sztucznej inteligencji w perspektywie przemian kultury i komunikacji **177**
DOI 10.12887/36-2023-3-143-11

WIZJE PRZYSZŁOŚCI

- * Jakub GOMUŁKA, Artificial Intelligence Applied to Philosophy: A Contribution to the Wittgenstein Ontology Project **207**
DOI 10.12887/36-2023-3-143-12
- * Leszek KORPOROWICZ – Redukcyjne pułapki socjologii „myślących maszyn” **235**
DOI 10.12887/36-2023-3-143-13
- * Renata TOMASZEWSKA – Raj – piekło – triumf. Sztuczna inteligencja w scenariuszach przyszłości **253**
DOI 10.12887/36-2023-3-143-14

MYŚLAĆ OJCZYŻNA...

- * Marek BERNACKI – „Czas przyszły jedną ogarnie nas nazwą / Niedoskonałych i już zapomnianych”. O nieznanym i zapomnianym wierszu Czesława Miłosza **277**
DOI 10.12887/36-2023-3-143-15

OMÓWIENIA I RECENZJE

- * Robert KUBLIKOWSKI – Filozofia sztucznej inteligencji (rec. M.A. Boden, *Sztuczna inteligencja. Jej natura i przyszłość. Krótkie wprowadzenie*, tłum. T. Sieczkowski, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2020) **293**
- * Propozycje „Ethosu” (J. Landgrebe, B. Smith, *Why Machines Will Never Rule the World: Artificial Intelligence without Fear*, Routledge, New York – London 2022) **298**

PRZEZ PRYZMAT ETHOSU

* Mirosława CHUDA – Oczekiwanie **303**

* Noty o autorach **309**

* Contents **315**

OD REDAKCJI

LEK I NADZIEJA

Systemy sztucznej inteligencji szybko wypełniają nasze bliższe i dalsze otoczenie i przejmują wiele funkcji wcześniejszych wytworów cywilizacji, oferując także nowe możliwości działania. Chociaż w otaczających nas urządzeniach elektronicznych fizycznie nie dostrzegamy obecności sztucznej inteligencji, to właśnie ona ustala naszą lokalizację, słucha i odpowiada na pytania, jakie zadajemy, „poszerza” nasze zdolności komunikowania, skraca czas, który poświęcamy na poszukiwanie nowych i ważnych dla nas informacji. Sztuczna inteligencja staje się też w coraz większym stopniu uczestnikiem życia społecznego. Humanoidalne roboty zaczynają sprawować opiekę nad osobami starszymi¹, a tak zwane roboty empatyczne wspomagają edukację dzieci i stają się ich towarzyszami². Jednym z najsłynniejszych robotów humanoidalnych jest Sophia (ze względu na podobieństwo do kobiety zwana fembotem), która w październiku 2017 roku otrzymała obywatelstwo Arabii Saudyjskiej, występowała na forum Organizacji Narodów Zjednoczonych, a obecnie szuka męża³. W czerwcu 2022 roku pracownik Google, Blake Lemoine, stwierdził, że wykorzystujący sztuczną inteligencję chatbot o nazwie LaMDA zyskał świadomość i zdolność odczuwania emocji (Google zaprzeczył, a chatbot wynajął prawnika...) ⁴. Z wielką ostrością pojawiają się przed nami ważne kwestie: Czy sztuczna inteligencja jest kolejnym etapem rozwoju

¹ Zob. M. Andtfolk, L. Nyholm, H. Eide, L. Fagerström, *Humanoid Robots in the Care of Older Persons: A Scoping Review*, „Assistive Technology” 34(2022) nr 5, s. 518-526.

² Zob. I. Leite, G. Castellano, A. Pereira in., *Empathic Robots for Long-term Interaction*, „International Journal of Social Robotics” 6(2014) nr 3, s. 329-341.

³ *Sophia ma obywatelstwo Arabii Saudyjskiej i szuka męża. Jest robotem*, dobreprogramy, <https://www.dobreprogramy.pl/sophia-ma-obywatelstwo-arabii-saudyjskiej-i-szuka-meza-jest-robotem,6820095894465376a>

⁴ Zob. J. Küveler, *Sztucznej inteligencji jest smutno. LaMDA przekonuje, że jest „osobą”, ze świadomością i życiem wewnętrznym*, Onet.pl, <https://kultura.onet.pl/lamda-sztuczna-inteligencja-google-twierdzi-ze-jest-swiadoma/htclckj>.

ludzkiej wolności i sprawczości, czy ich zagrożeniem? Czy rozwój sztucznej inteligencji zmierza do zbudowania bytów obdarzonych świadomością, intelektem i wolną wolą? Czy tworzenie sztucznej inteligencji uczy nas czegoś o naszych własnych umysłach i – wobec tego – pozwoli na ich ulepszanie? Pytania można mnożyć, ale jedno nie ulega wątpliwości: sztuczna inteligencja pozostanie z nami, a zatem nie mamy wyjścia, musimy ją – i się z nią – oswoić, jeśli nie chcemy stawać się coraz bardziej bezradni i niepewni w obliczu jej nieuchronnego oddziaływania na nasze życie.

Na czym takie oswojenie miałyby polegać? Sztuczna inteligencja tak szybko weszła na rynek, że nawet nie zdajemy sobie sprawy, w ilu używanych przez nas urządzeniach jest obecna. W tym sensie się z nią oswajamy: przyzwyczajamy się do niej i zaczynamy ją traktować jako (już) oczywisty i neutralny element naszego świata. Czy jednak rzecz jest aż tak prosta? Wydaje się, że nie, a pewne intuicje potwierdzające ten fakt znajdziemy w Słowniku Języka Polskiego PWN: „«Oswoić» to 1. «przyzwyczaić do kogoś lub do czegoś albo zapoznać z czymś»; 2. «przyzwyczaić dzikie zwierzęta do przebywania wśród ludzi lub do służenia ludziom»”⁵. Pierwsze znaczenie sugeruje, że oswajanie sztucznej inteligencji wymaga nabywania wiedzy na jej temat. „Oswoić” znaczy tu do pewnego stopnia „zrozumieć”. Problemami dotyczącymi sztucznej inteligencji z pewnością zajmują się przedstawiciele nauk ścisłych i technicznych, ale jej poznanie ma też inny aspekt: podobnie jak wszelkie artefakty, wymaga ona szerokiego namysłu filozofów, kulturoznawców, teoretyków informacji i komunikologów, socjologów, pedagogów, przedstawicieli nauk humanistycznych, a nawet teologów. Porzucamy bowiem powoli ideę, że artefakty techniczne stanowią jedynie przedłużenie ludzkiej ręki, umysłu i woli. Są one nasycone wartościami, a ich zaistnienie i funkcjonowanie zmienia nasz świat i nas samych⁶. Badania nad sztuczną inteligencją i jej sukcesy wybitnie przyczyniły się do rozpoznania tego faktu, powodując burzliwe dyskusje z ogromną polaryzacją stanowisk, której towarzyszą równie silne emocje – tak pozytywne, jak i negatywne. Trudno spodziewać się, by „zwykły zjadacz chleba” był w stanie choć pobieżnie opanować wiedzę specjalistyczną i zrozumieć partykularne rozwiązania informatyczne, ale już namysł nad konsekwencjami funkcjonowania i wykorzystywania sztucznej inteligencji staje się naszym obowiązkiem właśnie jako jej użytkowników. Wszak decydując się na przykład na wzięcie kota, powinniśmy się nauczyć co nieco o kotach, o ich psychice, o konsekwencjach posiadania kota, o możliwych zmianach w domowych relacjach... Per analogiam można to samo powiedzieć

⁵ Słownik Języka Polskiego PWN, hasło „Oswoić”, <https://sjp.pwn.pl/szukaj/oswoi%C4%87.html>.

⁶ Zob. R. L i z u t, *Technika a wartości. Spór o aksjologiczną neutralność artefaktów*, Wydawnictwo Naukowe Academicon, Lublin 2014

o sztucznej inteligencji, nawet jeśli nie jest to – jeszcze – stworzenie żywe. Ważne jest przede wszystkim spojrzenie na relację człowiek–sztuczna inteligencja. Rysują się dwa istotne aspekty tej relacji. Pierwszy wiąże się z możliwością tworzenia robotów humanoidalnych czy systemów informatycznych przyjaznych bądź nieprzyjaznych człowiekowi. W tym właśnie kontekście pojawia się sugestia implikowana przez drugie słownikowe znaczenie terminu „oswoić”: przystosować „do służenia ludziom”. Postulat taki został wprost wyrażony w roku 2017 podczas zorganizowanej przez Future of Life Institute konferencji „Beneficial AI 2017”, która odbyła się w Asilomar w Kalifornii i zgromadziła twórców sztucznej inteligencji z całego świata. Sformułowano wówczas zasady, którymi należy się kierować rozwijając sztuczną inteligencję. Według pierwszej z nich „celem badań nad rozwijaniem SI nie powinno być tworzenie inteligencji nieukierunkowanej, ale inteligencji dobroczynnej”⁷; zasada jedenasta głosi natomiast, że „systemy SI powinny być tak projektowane i tak działać, by były kompatybilne z ideałami godności, praw i wolności człowieka oraz kulturowej różnorodności”⁸. Ostatnia zaś, dwudziesta trzecia zasada, dotycząca rozwoju tak zwanej mocnej sztucznej inteligencji, podkreśla: „Superinteligencja powinna być rozwijana jedynie w służbie powszechnie wyznawanych ideałów etycznych i służyć dobru ludzkości, nie zaś jednemu państwu czy jednej organizacji”⁹. Łatwiej oczywiście powiedzieć niż zrobić. I tu pojawia się drugi ważny aspekt relacji człowiek–sztuczna inteligencja, dotyczący algorytmów, dzięki którym sztuczna oddziałuje na praktyki społeczne i organizacyjne grup ludzkich, przede wszystkim na strategię komunikacyjną, sposoby określania własnej tożsamości, modele tworzenia wspólnot przy użyciu pośredników cyfrowych czy sposoby sprawowania władzy i kontroli w organizacjach. Warto w tym kontekście rozważyć także, w jaki sposób zmienia się nasze życie, gdy wyposażone w sztuczną inteligencję maszyny przestaną być jedynie narzędziami i staną się integralnymi częściami naszych ciał i umysłów. Z jednej strony wydaje się, że nie chcemy utracić „siebie”: świadomości i odrębności własnego istnienia, ale z drugiej strony chcemy żyć wygodnie i bezpiecznie, zachowując jakość życia odpowiednią dla obowiązującego wzorca kulturowego. Choć sztuczna inteligencja nie jest jeszcze istotą samodzielnie myślącą, ludzkie myślenie o świecie i sam ludzki świat niewątpliwie zmieniają się w następstwie jej wykorzystania przez człowieka. Realizując pragnienie łatwości i szybkości dostępu do informacji, stajemy się jednocześnie „widoczni” dla cyfrowych oczu. Jak wówczas „wyważyć” granicę między korzyściami

⁷ *AI Principles*, Future of Life Institute, <https://futureoflife.org/open-letter/ai-principles/>. Tłumaczenie fragmentów obcojęzycznych – A.L.K.

⁸ Tamże.

⁹ Tamże.

wynikającymi z korzystania ze sztucznej inteligencji, a niebezpieczeństwem utraty prywatności? Obecnie to nie tylko dylemat użytkownika narzędzi elektronicznych, ale także jeden z głównych problemów cywilizacyjnych.

Proces „oswajania” sztucznej inteligencji zmienia człowieka tak, jak udomowienie zwierząt zmieniło życie ówczesnych łowców. W obrębie naszego świata i akceptowanych przez nas wartości, takich jak wymiana, współpraca czy komunikacja, musimy wykształcić w sobie nowe umiejętności, a być może nawet przededefiniować istotę rozmaitych procesów czy relacji międzyludzkich, jak choćby przyjaźń, współpraca czy troska o jakość życia. Analogia między procesem udomawiania zwierząt a osławianiem sztucznej inteligencji nie jest jednak całkowita. Jako twórcy inteligentnych maszyn mamy bowiem trudność z podjęciem decyzji, czy chodzi nam o coraz doskonalsze urządzenia, czy też o stworzenie istot autonomicznych, które będą podejmować decyzje niejako bez nas. Ta druga ewentualność stała się już w pewnym stopniu rzeczywistością. W roku 2021 głośnym echem odbił się raport Rady Bezpieczeństwa ONZ do spraw Libii na temat dronów bojowych zastosowanych w w tym kraju, które same namierzyły i zaatakowały żołnierzy (i prawdopodobnie zabiły) bez rozkazu człowieka¹⁰. Istnieją też tak zwane roboty społeczne, które Kate Darling z Massachusetts Institute of Technology definiuje jako materialne, autonomiczne podmioty działające, komunikujące się i wchodzące w interakcje z człowiekiem na poziomie emocjonalnym¹¹. Są one czymś więcej niż roboty autonomiczne, realizujące zadania niezależnie i w znacznych odległościach od człowieka. To roboty, które mają wchodzić w interakcje i współpracować z człowiekiem jako jego partnerzy, a być może nawet zastępować ludzi jako naturalnych partnerów komunikacji. Patrząc na osiągnięcia robotów empatycznych, jakim jest na przykład KASPAR (Kinesics and Synchronization in Personal Assistant Robotics), służących dzieciom autystycznym jako towarzysze¹², czy robotów w rodzaju Paro czy Pearl, zaprojektowanych jako wspomaganie osób starszych w ich codzienności¹³, przestajemy się dziwić, gdy docierają do nas informacje o badaniach nad psychologicznym kontraktem między człowiekiem a robotem¹⁴, o pojawieniu

¹⁰ Zob. A. Stryewska, *Autonomiczny dron zaatakował żołnierzy*, PCWorld, <https://www.pcworld.pl/news/Autonomiczny-dron-zaatakowal-zolnierzy,428254.html>.

¹¹ Por. K. Darling, *Extending Legal Protection to Social Robots*, w: *Robot Law*, red. R. Calo, M. Fromkin, I. Kerr, Edward Elgar Publishing, Cheltenham, UK, s. 215.

¹² Zob. *Kaspar the Social Robot*, University of Hertfordshire, <https://www.herts.ac.uk/kaspar/the-social-robot>.

¹³ Zob. J. Broekens i in., *Assistive Social Robots in Elderly Care: A Review*, „Gerontechnology” 8(2009), nr 2, s. 94-103.

¹⁴ Zob. np. A. Rogozińska-Pawełczyk, *Towards Discovering Employee-Robot Interaction: Aspects of Concluding the Psychological Contract*, „Education of Economists and Managers” 58(2020) nr 4, s. 9-20. Kontrakt psychologiczny to zbiór wzajemnych niepisanych oczekiwań istniejących między pracownikami a pracodawcami w organizacji.

się psychiatrii robotów¹⁵ czy też o przyznawaniu praw robotom i nadawaniu im osobowości prawnej¹⁶. Być może zatem rację miał lis z *Małego księcia* Antoine’a de Saint-Exupéry’ego, twierdząc, że „oswoić” znaczy „stworzyć więzy”¹⁷? Lis wyjaśniał jeszcze dokładniej: „Jeżeli mnie oswoisz, będziemy się nawzajem potrzebować. Będziesz dla mnie jedyny na świecie. I ja będę dla ciebie jedyny na świecie”¹⁸. Czy będziemy zatem kiedyś mówić: „Mój ty jedyny, najdroższy robociku”? Badacze pracujący w nowej dziedzinie HRI (human-robot interaction) twierdzą, że już to się do pewnego stopnia dzieje i nasze emocjonalne podejście do robotów zmienia się radykalnie: ufamy ich radom, przywiązujemy się do nich, a nawet traktujemy jako życiowych partnerów¹⁹.

Rozwój sztucznej inteligencji nas fascynuje, a zarazem przeraża. Jest w nas strach Wiktora Frankensteina. Chcąc stworzyć „sztuczną” istotę, obdarzoną dużą inteligencją i wykonującą polecenia stwórcy, powołał do istnienia potwora, który zamiast pomagać, stał się dla ludzi zagrożeniem. Historia jest wprawdzie fantastyczna, ale lęk przed „sztucznym” człowiekiem jest jak najbardziej realny. Młoda angielska pisarka Mary Shelley zatytułowała swą powieść z 1818 roku *Frankenstein, czyli Współczesny Prometeusz*²⁰. Zestawienie „wytworu” doktora Frankensteina ze „współczesnym Prometeuszem” jest zaskakujące, a jednocześnie trafne. Mitologiczny Prometeusz ulepił i ożywił człowieka, wbrew woli Zeusa podarował mu ogień, a potem nauczył go przetapiać metale, uprawiać rolę, kuć zbroje, budować domy, czytać, pisać i ujarzmić siły przyrody. Dar życia i ognia wiązał się też jednak z zagrożeniami – z nieprzychylnością greckich bogów i karą, jaka spotkała człowieka w wyniku otwarcia przez Pandorę puszek z nieszczęściami. Frankenstein także tworzy człowieka – człowieka sztucznego, którego wyposaża w siłę i inteligencję, ale nie w rozumienie świata. Odrzucony przez ludzi, potwór obraca się przeciwko swojemu stwórcy i staje się niebezpieczny. Dziś to my stajemy się współczesnymi prometeuszami. Dzięki programistom, inżynierom, trenerom umiejętności sztucznej inteligencji otrzymujemy nowe technologie, które „rozszerzają” ludzkie możliwości. Otrzymujemy też inteligentne robo-

¹⁵ Zob. np. O. P a c e l t, *Psychiatra dla robota?* Joanne Pransky, <https://botland.com.pl/blog/psychiatra-dla-robota-joanne-pransky/>. Także W. Żółcińska, *Psychiatra robotów*, „Computerworld”, 27.08.2020, <https://www.computerworld.pl/wywiad/Psychiatra-robotow,422594.html>.

¹⁶ Zob. np. K. B i c z y s k o - P u d e ł k o, D. S z o s t e k, *Koncepcje dotyczące osobowości prawnej robotów – zagadnienia wybrane*, „Prawo Mediów Elektronicznych” 2019, nr 2, s. 9-15.

¹⁷ A. d e S a i n t - E x u p é r y, *Mały Książę*, tłum. J. Szwykowski, Instytut Wydawniczy Pax, Warszawa 1958, s. 61.

¹⁸ Tamże, s. 62.

¹⁹ Zob. np. *Dla gumowej lalki porzucił żonę i dzieci*, Polish Express. Styl życia, <https://www.polishexpress.co.uk/dla-gumowej-lalki-porzucil-zone-i-dzieci>.

²⁰ Zob. M. S h e l l e y, *Frankenstein, czyli Współczesny Prometeusz*, tłum. M. Piłża, Wydawnictwo Vesper, Czerwonak 2013.

ty, które wykonują za człowieka pracę. Dzięki sztucznej inteligencji mamy nowe możliwości wyszukiwania, segregowania i porządkowania nieprzebranej ilości danych; sztuczna inteligencja może pisać teksty, tworzyć obrazy, pisać pogromy komputerowe, rozpoznawać twarze i odpowiadać na nasze emocje. Coraz częściej w operacjach biorą udział roboty kardiochirurgiczne. Nanoboty umożliwiają diagnozowanie chorób i ich leczenie, pomagają w poszukiwaniu informacji o dawcach i biorcach organów o odpowiedniej zgodności tkankowej. Interfejs łączący ludzki mózg z komputerem może spowolnić procesy biologiczne związane z rozwojem chorób Parkinsona i Alzheimer, a także poprawiać naturalne ludzkie zdolności poznawcze²¹. To ewidentnie szanse dla poszczególnych ludzi, a być może dla całej ludzkości. Długo by można wymienić już osiągnięte i spodziewane korzyści rozwoju sztucznej inteligencji. Mity i literatura ostrzegają nas jednak przed przekraczaniem pewnych granic związanych z wykorzystywaniem techniki w odniesieniu do ludzkiego ciała i umysłu, nasze twory mogą bowiem „wyrwać się” spod naszej kontroli. Poważnie trzeba traktować pytanie, na jakie ingerencje w ludzkie ciało i umysł możemy sobie pozwolić, a jakie działania powinny być zabronione. Jakiego kierunku badań nad sztuczną inteligencją powinny być rozwijane, a jakie być należy porzucić? Pytania tego rodzaju wpisują się w perspektywę techniczną, ale też prawną, filozoficzno-etyczną, a być może również teologiczną²².

Proces osławiania natury, czyli udomowienia „dzikich” roślin i zwierząt, zabrał człowiekowi kilka tysięcy lat i miał swoje konsekwencje na przykład w postaci transferu chorób zakaźnych przenoszonych na ludzi. Obecnie nie mamy tak dużo czasu, musimy się szybciej przystosować do życia w świecie systemów techniki. Jakiego kosztu będzie miało wykorzystanie robotów w obszarze codziennego funkcjonowania oraz wprowadzenie elementów robotycznych w obręb ludzkich ciał i umysłów, by zwiększyć pojemność pamięci naszego mózgu i naszą zdolność przyswajania danych, by poprawić lub zwiększyć możliwości szybkiego poruszania się, percypowania rzeczy, przekazywania emocji na odległość? W obu wypadkach mogą zadziałać mechanizmy przystosowania, ale czy przystosowanie zawsze oznacza lepsze życie? Konsekwencją osławiania zwierząt stała się odporność biologiczna populacji ludzkiej, konsekwencją osławiania sztucznej inteligencji może być wytworzenie przez człowieka nowych kompetencji – informatycznych, technicznych, medialnych, ale też wymuszenie zmiany stylu jego życia. Ale czy tylko to?

²¹ Zob. K. Halićka, D. Surel, *Smart Living Technologies in the Context of Improving the Quality of Life for Older People: The Case of the Humanoid Rudy Robot*, „Human Technology” 18(2022) nr 2, s. 191-208.

²² Zob. E. Nosarzewska, 2030. *Czy sztuczna inteligencja może być etyczna? Raport PEW Reserch Center, Polskie Towarzystwo Studiów nad Przyszłością*, <https://ptspl/sztuczna-inteligencja-etyka/>.

Z jednej strony, wykorzystując sztuczną inteligencję, dostrzegamy szansę na „ulepszenie” ludzkiego życia, nie chcąc zarazem tracić z oczu związanych z tym zagrożeń. Chodzi przecież nie tylko o wydłużone świadome istnienie, ale także o życie sensowne. Z drugiej strony obawiamy się utraty wolności i świata, który poddany zostanie całkowitej kontroli i w którym każdy nasz ruch będzie drobiazgowo odnotowywany. W świecie takim utraty wolności możemy jednak nie zauważyć, współczesne formy zniewolenia mają bowiem postać „miękką” i dlatego tak trudno dostrzec, że niosą w sobie potencjalność zagrożenia. Pozostaje pytanie, co przeważy i czy jesteśmy gotowi, żeby zmierzyć się z wyzwaniami, które stają przed nami w związku z postępującym rozwojem sztucznej inteligencji? I czy ostatecznie sztuczna inteligencja się „usamodzielni”, uznając nas za niższy, a być może nawet zbędny i szkodliwy gatunek?

Pytania te stawiane są tak przez naukowców, jak i przez zwykłych ludzi, a doniesienia medialne o kolejnych sukcesach sztucznej inteligencji jedynie podsycają angażujące emocjonalnie spory. Emocje często jednak zaciemniają merytoryczną stronę problemu i dlatego niezbędne są platformy kompetentnych, rzetelnych, spokojnych debat nad kierunkiem rozwoju sztucznej inteligencji. Niniejszy tom „Ethosu” wychodzi naprzeciw tej potrzebie. Artykuły w nim zebrane tworzą interdyscyplinarną mapę problemów ważnych nie tylko z punktu widzenia środowisk zainteresowanych instrumentalnym wykorzystywaniem rosnących zasobów sztucznej inteligencji, ale także z punktu widzenia „zwykłego” użytkownika nowoczesnej techniki. Niezależnie bowiem od tego, jak bardzo dystansujemy się od współczesnych wytworów techniki, i tak zostajemy niejako „wkomponowani” w obszar jej oddziaływania. Autorzy artykułów analizują zjawiska kulturowe związane z funkcjonowaniem człowieka w coraz ściślejszej relacji z maszynami (robotami, chatbotami czy autonomicznymi systemami elektronicznymi). Interesują ich zarówno poziom ontologiczny zjawisk (czym w istocie jest coś, co nazywamy sztuczną inteligencją i jaki charakter bytowy mają relacje między człowiekiem a dynamicznie rozwijającą się jej rzeczywistością), jak i przemiany zachodzące w zbiorowej i indywidualnej świadomości człowieka, dokonujące się pod wpływem coraz intensywniejszego kontaktu z wytworami technologii informacyjnych. Ostatecznym celem jest więc analiza zjawisk związanych z wykorzystywaniem artefaktów świata technicznego, a być może nawet ze współbyciem człowieka z tego rodzaju artefaktami (poprzez rozszerzenie, cyborgizację czy augmentację). Namysł nad wszystkimi tymi zjawiskami stanowi element osvajania sztucznej inteligencji w sensie „zapoznania się” z nią oraz próby jej „zrozumienia”, a poprzez to niejako „uczynienia własnym” świata, w którym jest ona w coraz mniejszym stopniu instrumentem, a w coraz większym stopniu współtowarzyszem. Liczymy, że tom „Ethosu” poświęcony tym kwestiom stanowić będzie zachętę do

refleksji w czasach, gdy nadal towarzyszy nam nadzieja, że nie jest jeszcze za późno na świadome sterowanie kierunkiem prac nad sztuczną inteligencją.

Agnieszka Lekka-Kowalik
Mariusz Wojewoda

FROM THE EDITORS

FEAR AND HOPE

AI systems gradually assume control over our immediate as well as distant environments, embracing functions traditionally attributed to inventions that historically preceded them. Much as we are not aware of the presence of AI in the electronic devices we use, it is precisely AI that shows our location, responds to our inquiries, “broadens” our communication skills and capacities, as well as shortens the time we need to find new or important information. Moreover, AI has become to ever greater extent a participant in social life. Humanoid robots can now take care of the elderly,¹ and the empathetic ones are used as assistance in the education of children, not infrequently becoming their companions.² Among the most popular humanoid robots is Sophia, called a fembot, since it physically resembles a woman. Sophia was granted citizenship by Saudi Arabia in 2017 and is famous for having joined a United Nations meeting on artificial intelligence and sustainable development. Sophia has also expressed the need to have a baby.³ In June 2022, Google engineer Blake Lamoine claimed that the artificially intelligent chatbot generator LaMDA had developed consciousness and sentience (Google denied the fact, Lamoine was fired, and LaMDA hired a lawyer...).⁴ All these facts, as well as many other similar ones, prompt the need to address the resulting grave issues: Is AI merely a new stage in the development of human agency and freedom, or does it pose a threat to them? Is AI development bound to produce beings endowed with consciousness, intellect, and free will? Does AI development teach us something about our own minds, and will it, consequently, help us

¹ See Malin A nd t f o l k, Linda N y h o l m, Hilde E i d e, Lisbeth F a g e r s t r ö m, “Humanoid Robots in the Care of Older Persons: A Scoping Review,” *Assistive Technology* 34, no. 5 (2022): 518–26.

² See Iolanda L e i t e, Ginevra C a s t e l l a n o, André P e r e i r a, et al., “Empathic Robots for Long-term Interaction,” *International Journal of Social Robotics* 6, no. 3 (2014): 329–41.

³ “Sophia the Robot Wants a Baby and Says Family Is ‘Really Important,’” BBC News, <https://www.bbc.com/news/newsbeat-42122742>.

⁴ See Margaret D a v i s, “Sentient AI LaMDA Hired a Lawyer to Advocate for Its Rights ‘As a Person,’ Google Engineer Claims,” *The Science Times*, <https://www.sciencetimes.com/articles/38379/20220625/sentient-ai-lamda-hired-lawyer-advocate-rights-person-google-engineer.htm>.

enhance them? While questions of this kind can be multiplied, one thing is beyond doubt: we will have to live with AI, and we have no other option but to tame—domesticate—it, become accustomed to its presence, unless we choose a passive stance towards the impact it is bound to make on our lives.

However, in what might ‘domestication’ of AI consist? The fact is that AI has entered the global market so fast that we do not even realize in how many devices we constantly use it is actually implemented. One might say that in this sense we have already begun domesticating AI, or, indeed, more than that: we already consider it as an obvious and neutral presence within our human universe. And yet is the issue really that simple? A mere look into the definition of “domesticate” suggests otherwise. As synonyms for “domesticate,” Thesaurus.com suggests “familiarize” or “accustom,” which hints at the fact that domestication in a way involves an understanding of its object or getting acquainted with it.⁵ This in turn suggests that domestication presupposes knowing the object to be domesticated. The Merriam-Webster Dictionary, interestingly, gives the following definition of the adjective “tame,” which it lists as a synonym for “domesticated”: “Reduced from a state of native wildness especially so as to be tractable and useful to humans.”⁶ Let us explore these insights.

Thus domesticating AI presupposes constant exploration of ways in which it is created and of the modes of its functioning. In this sense “domesticating” means “understanding.” While technological problems related to AI will be certainly addressed by IT specialists and communicologists, understanding it embraces also research into aspects which go beyond technology-based issues. As it is the case with any other artifact, AI needs to be systematically studied by philosophers as well as by scholars in fields such as cultural studies, sociology, education, the humanities, and, last but not least, theology. We are gradually departing from the idea that technical artifacts are merely extensions of the human hand, mind or will. Indeed, they are value-laden, and their existence and constant presence affects the universe around us as much as it affects ourselves.⁷ The ongoing research into AI and its success has substantially contributed to the recognition of this fact, not infrequently inciting heated debates in which strongly polarized standpoints are adopted and which stir up strong, positive as well as negative, emotions. While we can hardly expect an ordinary person to have even a loose grasp of the expertise involved in AI creation, or to comprehend particular IT solutions, insight into the consequences of AI and its applications is our duty, since it is us who use it. One can venture to bring up an

⁵ See Thesaurus.com, Synonyms for “domesticate,” <https://www.thesaurus.com/browse/domesticate>.

⁶ Merriam-Webster Dictionary, s.v. “Tame,” <https://www.merriam-webster.com/dictionary/tame>.

⁷ See Rafał L i z u t, *Technika a wartości: Spór o aksjologiczną neutralność artefaktów*, Lublin: Wydawnictwo Naukowe Academicon, 2014

analogy: Once we decide to adopt a cat, we must learn as much as we can about cats, their behavior and needs, as well as about the change in family relations a cat may trigger. Despite the fact that AI is not (yet) a living creature, adopting it should be approached in a similar way. What comes to the foreground in this context is the relationship between AI and humans, which has two significant aspects. The first one is the possibility of building humanoid robots (or IT systems) which can be either friendly or unfriendly to human beings. It is in this context that one of the meanings of “domestication” The Merriam-Webster Dictionary lists becomes important: “domesticate” means here “adapt something to be beneficial for human beings.” A postulate of exactly this kind was included among the “AI Principles” proposed at the 2017 Asilomar Conference “Beneficial AI 2017,” which gathered AI researchers from all over the world. Principle one stated: “The goal of AI research should be to create not undirected intelligence, but beneficial intelligence”⁸; principle eleven held: “AI systems should be designed and operated so as to be compatible with ideals of human dignity, rights, freedoms, and cultural diversity,”⁹ and principle twenty three underscored: “Superintelligence should only be developed in the service of widely shared ethical ideals, and for the benefit of all humanity rather than one state or organization,”¹⁰ which is certainly easier said than done. And it is here that the second important aspect of the relation between the human beings and AI surfaces, namely, that of the algorithms which make it possible to affect social and organizational practices within human communities, such as communication strategies, ways of defining identity, community building models, and ways of executing power and control within organizations. In this context, it also worthwhile considering the potential change in our lives which will take place once AI endowed machines stop being merely tools and become integral elements of our minds and bodies. On the one hand, it seems that we do not wish to be rid of “ourselves”: our consciousnesses and separate existences, yet, on the other hand, we want to live comfortably, enjoying safety and the quality of life adequate to the current cultural pattern. And even though AI is currently incapable of thinking, the ways humans think about the world, as well as the world itself, are changing due to their use of AI. While satisfying our need for an easy and quick access to information, we have become visible to the “digital eyes.” How then to balance the benefits of recourse to AI and the risk of losing privacy our recourse to AI involves? Nowadays, such a dilemma goes beyond the choice an individual IT user must make; indeed, it needs to be considered among the crucial civilizational issues.

⁸ “AI Principles,” Future of Life Institute, <https://futureoflife.org/open-letter/ai-principles/>.

⁹ Ibidem.

¹⁰ Ibidem.

The process of “domesticating” AI has changed human beings, just as domestication of animals once changed the lives of hunter-gatherers. In our human universe, in which values such as exchange, cooperation or communication are commonly accepted, we now need to develop new skills, or even redefine the essence of various social processes and interpersonal relationships, for instance, those of friendship and cooperation, as well as our concern for the quality of life. However, the analogy between the process of the domestication of animals and that of “taming” AI is not entirely adequate. Creators of intelligent machines, we have difficulty deciding whether we wish to build just more and more perfect devices, or maybe autonomous beings capable of making their own decisions. The second option in a sense already came true. In 2021, a United Nations report about a March 2020 skirmish in the Libyan military conflict reverberated around the world. It revealed that a drone, a lethal autonomous weapons system, may have aimed at, attacked (and probably killed) soldiers without being instructed to do so by a human being.¹¹ One can think in this context also about the so-called social robots Kate Darling from the Massachusetts Institute of Technology defines as materially incarnate, autonomous actors that communicate and interact with humans on an emotional level.¹² They are something more than autonomous robots which perform tasks independently and at considerable distances from humans. These are robots that should interact and collaborate with human beings as partners, and maybe even replace human beings as natural partners in communication. When we consider the achievements of such empathetic robots as KASPAR (Kinesics and Synchronization in Personal Assistant Robotics) that accompany autistic children,¹³ or robots such as Paro or Pearl, designed to help the elderly in their daily activities,¹⁴ we will probably not be surprised by the information about research on the psychological contract between robot and human¹⁵, about de-

¹¹ See Dave M a k i c h u k, “Autonomous Drone May Have Killed in Libya: UN Report,” Asia Times, <https://asiatimes.com/2021/06/autonomous-drone-may-have-killed-in-libya-un-report/>.

¹² See Kate D a r l i n g, *Extending Legal Protection to Social Robots*, in: *Robot Law*, ed. Ryan Calo, Michael Froomkin, and Ian Kerr, Cheltenham, UK: Edward Elgar Publishing, 215.

¹³ See *Kaspar the Social Robot*, University of Hertfordshire, <https://www.herts.ac.uk/kaspar/the-social-robot>.

¹⁴ See Joost B r o e k e n s et al., “Assistive Social Robots in Elderly Care: A Review,” *Gerontechnology* 8, no. 2 (2009): 94–103.

¹⁵ See, e.g., Anna R o g o z i ń s k a - P a w e ł c z y k, “Towards Discovering Employee-Robot Interaction: Aspects of Concluding the Psychological Contract,” *Education of Economists and Managers* 58, no. 4 (2020): 9–20. A psychological contract is “the unwritten, intangible agreement between an employee and their employer that describes the informal commitments, expectations and understandings that make up their relationship.” Jack E n r i g h t, “What Is the Psychological Contract?” Charlie Makes HR Effortless, <https://www.charliehr.com/blog/what-is-the-psychological-contract/>.

veloping AI psychiatry,¹⁶ or about granting robots certain rights and making them legal personalities.¹⁷ So perhaps we can say that the fox in Antoine de Saint-Exupéry's *Little Prince* was right, when he said that "tame" means "establish ties."¹⁸ The fox explained further: "If you tame me, then we shall need each other. To me, you will be unique in all the world. To you I shall be unique in all the world."¹⁹ So, will the time come for us to address our robots, saying things such as: "My unique, dearest little robot"? Researchers working in the new field called Human–Robot Interaction (HRI) claim this is already happening, at least to a certain extent, for our emotional attitudes towards robots are changing radically: we trust their advice, we get attached to them, and even choose them as life partners.²⁰

AI development both fascinates and frightens us. In a way, we harbor the kind of fear Victor Frankenstein must have experienced. He merely wanted to create an "artificial" human being, endowed with a high intelligence and acting on the commands of its creator, but what he brought into existence turned out a monster, a threat rather than help to people. Although the story in question is a piece of fiction, the fear of an "artificial" human being is no less than real. In 1818, Mary Shelley, a young British writer, entitled her novel *Frankenstein or The Modern Prometheus*.²¹ Astonishing as it might seem, juxtaposing the artifact Victor Frankenstein produced with "the modern Prometheus" is absolutely right. Mythological Prometheus created man from clay and gave him life; he then stole fire from the gods and gave it to humanity, which enraged Zeus. Prometheus taught humans how to melt metals, farm, forge armor, build houses, read, write, and harness the forces of nature. However, the gift of life and fire was also associated with danger—with the disfavor of the Greek gods and the punishment man had to suffer once Pandora had opened the box with misfortunes. Frankenstein also created a man: an artificial man whom he equipped with physical strength and intelligence, yet one devoid of an understanding of the world. Rejected by people, the monster turned against its

¹⁶ See, e.g., Mark Sackler, "Joanne Pransky—World's First Robotic Psychiatrist," Age of Robots, <https://ageofrobots.net/joanne-pransky-worlds-first-robotic-psychiatrist/>.

¹⁷ See, e.g., Sergio Avila Negriko, "Robot as Legal Person: Electronic Personhood in Robotics and Artificial Intelligence," *Frontiers in Robotics and AI*, Dec. 2021, 8, https://www.researchgate.net/publication/357296496_Robot_as_Legal_Person_Electronic_Personhood_in_Robotics_and_Artificial_Intelligence.

¹⁸ Antoine de Saint-Exupéry, *The Little Prince*, trans. Richard Howard, Boston and New York: Harcourt, 2000, 59.

¹⁹ Ibidem.

²⁰ See, e.g., Leotronics.eu, "Marriage with a Robot: The Future of Humanity?" <https://leotronics.eu/en/blog/marriage-with-a-robot-the-future-of-humanity>.

²¹ See Mary Shelley, *Frankenstein or The Modern Prometheus*, London: Penguin Books, 2003.

creator and became dangerous. Nowadays, it is us who have become modern Prometheuses. Owing to the work continued by IT specialists, engineers, and trainers of AI learning, we have gained access to new technologies which have the potentiality of enhancing human potentialities. We get intelligent robots that perform tasks hitherto belonging to humans and we have new possibilities of accessing enormous amounts of data, sorting and filing them. AI is capable of generating human-like texts, creating images, writing computer programs, recognizing faces and responding to our emotions. Robotic systems have been successfully utilized in cardiac surgeries. Nanobots are used as a tool to diagnose illnesses and help deal with them; they are also useful for the purpose of searching for information about potential organ donors with proper tissue compatibility. A brain–computer interface is capable of slowing down certain biological processes characteristic of Parkinson’s and Alzheimer’s diseases, as well as of improving natural human cognitive abilities.²² Such advancements can serve individual persons as well as humanity. A list of the already accomplished and expected benefits of the development of AI would take a long time to make. However, myths as well as fiction warn us against crossing certain limits of applying *techné* to the human body and mind, since our creations can escape our control.

The question of which interventions in the human body and mind can be allowed and which must be prohibited needs to be taken seriously. What kind of research on AI should be developed, what kind of it must be abandoned? Such questions enter well not only into the technical perspective, but also into the legal, philosophical, ethical, and perhaps also theological ones.²³

The process of taming nature, i.e., domesticating “wild” plants and animals, was stretched over millennia, and it had various consequences: among others, some infectious diseases were transferred from animals to humans. Domesticating AI, however, must be performed at a much quicker pace, and we will soon need to adapt ourselves to living in the world of technological systems. What will be the price of using robots in everyday life and of introducing robotic elements into the human minds and bodies in order to enhance brain capacity, memory, and capability of data-mastering, energy needed to perform physical activities, perception, or sharing of emotions regardless of the physical distance? We can certainly count on human adaptation mechanisms,

²² See Katarzyna Halićka and Dariusz Surul, “Smart Living Technologies in the Context of Improving the Quality of Life for Older People: The Case of the Humanoid Rudy Robot,” *Human Technology* 18, no. 2 (2022): 191–208.

²³ See, e.g., Albert M. Erisman and Tripp Parker, “Artificial Intelligence: A Theological Perspective,” *Perspectives on Science and Christian Faith* 71, no. 2 (2019): 95–106; Stanford Encyclopedia of Philosophy, s.v. “Ethics of Artificial Intelligence and Robotics” (by Vincent C. Müller), <https://plato.stanford.edu/entries/ethics-ai/>.

but does being adapted to a new environment necessarily signify a better life? Among the consequences of domesticating animals was enhancement of the immune system of the human population, and it is true that the impact of the domestication of AI may involve developing new informational, technical, and media-related skills and competences by humans, as well as introducing changes in their lifestyles. However, will there be more to it? In utilizing AI, we see a chance to “enhance” human life as such, yet, we want to monitor the dangers the processes in question may trigger. The goal is not merely a prolonged life, but also a meaningful one. Still, we are afraid that, in the world which will be subject to total control and in which our every move will be monitored, we will be bound to lose our freedom. And yet, much as it seems a paradox, a loss of freedom may not be even discernible in the world to come, since modernity is characterized by “soft” enslavement, among others, and potential threats to freedom are by no means easy to define. The question remains of what will prevail and whether we are ready to meet the challenges we face as a result of AI continuous development. Will AI eventually become independent and begin to consider humans as a lower and maybe even harmful or useless species? Questions of this kind are posed by scientists as well as by ordinary people, and media reports only fuel emotionally engaging disputes. Emotions, however, often obscure the substantive side of the problem, which is why platforms for competent, reliable, and calm debates on the direction of the AI development are so necessary. The present volume of *Ethos* is intended as a response to this need. The articles we have collected outline a multidisciplinary map of AI-related problems which are important not only to those interested in the actual application of the growing resources of AI tools and services, but also to an ordinary user of modern technology. For regardless of how much we distance ourselves from the products of contemporary *techne*, we are in one way or another “incorporated” into the area of its influence. The authors analyze cultural phenomena related to the functioning of humans who are in an increasingly close relationship with machines (robots, chatbots or autonomous electronic systems). They are interested in both the ontological level of the phenomena (the essence of what we call AI and the existential nature of the relations between the human being and the dynamically developing reality of AI) and the changes occurring in the collective and individual consciousness under the influence of an increasingly intense contact with IT products. The ultimate goal is therefore to analyze phenomena related to the use of technological AI artifacts and perhaps even to the coexistence of humans with such artifacts (through enhancement, cyborgization or augmentation). Reflection on these phenomena is an element of taming artificial intelligence in the sense of “getting acquainted” with it and trying to “understand” it, and thus “making our own” the world in which AI is less and less an instrument, and

more and more a companion. We envision that the volume of *Ethos* dedicated to such issues will encourage reflection in this peculiar moment of time when we still hope that it is not too late to consciously steer the direction of work on artificial intelligence.

Agnieszka Lekka-Kowalik and Mariusz Wojewoda

JAN PAWEŁ II
SZYBKI ROZWÓJ

JOHN PAUL II
THE RAPID DEVELOPMENT

JAN PAWEŁ II

SZYBKI ROZWÓJ*

Szybki rozwój techniki w dziedzinie środków przekazu jest z pewnością jedną z oznak postępu współczesnego społeczeństwa. Gdy patrzymy na nieustanną ewolucję nowych wynalazków, jeszcze bardziej aktualne wydają się słowa Dekretu Powszechnego Soboru Watykańskiego II *Inter mirifica*, promulgowanego przez mego czcigodnego poprzednika sługę Bożego Pawła VI w dniu 4 grudnia 1963 roku: „Wśród niezwykłych wynalazków techniki, które zwłaszcza w najnowszych czasach ludzki geniusz z pomocą Bożą rozwinął spośród rzeczy stworzonych, Matka Kościół ze szczególnym zainteresowaniem przyjmuje i popiera te, które w największym stopniu dotyczą ducha ludzkiego i które otworzyły nowe drogi bardzo łatwego przekazywania wszelkiego rodzaju informacji, przemysłów i zaleceń”¹.

Dzisiaj, ponad czterdzieści lat po publikacji tego dokumentu, bardzo potrzebne wydaje się podjęcie na nowo refleksji nad «wyzwaniami», jakie stanowią dla Kościoła środki przekazu. Kościół – jak pisał Paweł VI – „byłby się winny przed swoim Panem, gdyby nie używał tych potężnych pomocy”². Kościół w istocie winien nie tylko posługiwać się mediami w celu szerzenia Ewangelii, ale – zwłaszcza dzisiaj – zespalać zbawcze orędzie z „nową kulturą”, którą potężne środki przekazu tworzą i rozpowszechniają. Kościół zdaje sobie sprawę, że stosowanie nowoczesnych technik i technologii przekazu społecznego jest integralną częścią jego misji w trzecim tysiącleciu.

Świadomość tego sprawia, że społeczność chrześcijańska poczyniła znaczne postępy w wykorzystywaniu środków społecznego przekazu w ta-

* Fragmenty Listu apostolskiego do odpowiedzialnych za środki społecznego przekazu (Watykan, 24 I 2005). Tekst publikujemy za: „L'Osservatore Romano” wyd. pol. 26(2005) nr 7-8(275), s. 35-39

¹ Sobór Watykański II, Dekret o środkach społecznego przekazu *Inter mirifica*, w: Sobór Watykański II, *Konstytucje. Dekrety. Deklaracje*, Pallottinum, Poznań 2002, s. 87.

² P a w e ł VI, Adhortacja apostolska *Evangelii nuntiandi*, nr 45, Opoka, https://opoka.org.pl/biblioteka/W/WP/pawel_vi/adhortacje/evangelii_nuntiandi.html.

kich dziedzinach, jak informacja religijna, ewangelizacja i katecheza, formacja duszpasterska do pracy w tym sektorze oraz wychowanie do dojrzałego i odpowiedzialnego korzystania z różnych środków społecznego przekazu.

Różne wyzwania stają przed nową ewangelizacją w świecie, w którym istnieją tak wielkie możliwości przekazu. Mając to na uwadze, w Encyklice *Redemptoris missio*³ podkreśliłem, że pierwszym areopagiem współczesnej epoki jest świat społecznego przekazu, który jest w stanie jednoczyć ludzkość i sprawić, że staje się – jak zwykło się mawiać – „światową wioską”⁴. Środki społecznego przekazu zyskały tak wielkie znaczenie, że dla wielu ludzi stały się głównym źródłem wskazań i inspiracji w sferze zachowań indywidualnych, rodzinnych i społecznych. Jest to problem złożony, ponieważ podłożem, z którego wyrasta tego rodzaju kultura, są nie tyle jakieś treści, ile raczej sam fakt istnienia nowych sposobów przekazu informacji, posługującego się nieznanymi dotąd technikami i językami. Żyjemy w epoce globalnego przekazu, w której wiele momentów ludzkiej egzystencji pozostaje pod wpływem procesów medialnych, a w każdym razie trzeba je uwzględniać. Wspomnę tu tylko o formowaniu osobowości i sumienia, sposobu rozumienia i budowania więzi uczuciowych, o kształtowaniu kolejnych etapów wychowania i formacji, o tworzeniu i upowszechnianiu zjawisk kulturowych, o rozwoju życia społecznego, politycznego i gospodarczego.

W organicznym i poprawnym modelu rozwoju człowieka środki przekazu mogą i powinny szerzyć sprawiedliwość i solidarność, przedstawiając wydarzenia w sposób ścisły i prawdziwy, analizując wnikliwie sytuacje i problemy, prezentując różne opinie. W dojrzałym i odpowiedzialnym korzystaniu z wolności nadrzędne kryteria prawdy i sprawiedliwości stanowią horyzont, w którym mieści się autentyczna deontologia, regulująca posługiwanie się nowoczesnymi potężnymi środkami przekazu.

Również świat mediów potrzebuje Chrystusowego odkupienia. W analizowaniu procesów zachodzących w dziedzinie środków przekazu i ich wartości w perspektywie wiary niewątpliwie pomocne może być zgłębianie Pisma Świętego. Jawi się ono bowiem jako „wielki kodeks” przekazu przesłania, które nie jest tylko chwilowe i okazjonalne, ale fundamentalne, ze względu na jego wymiar zbawczy.

Historia zbawienia opisuje i poświadcza komunikowanie się Boga z ludźmi; w tej komunikacji wykorzystywane są wszystkie formy i odmiany przekazu. Człowiek został stworzony na obraz i podobieństwo Boga, aby przyjął

³ Zob. Jan Paweł II, Encyklika *Redemptoris missio* o stałej aktualności posłania misyjnego, The Holy See, https://www.vatican.va/content/john-paul-ii/pl/encyclicals/documents/hf_jp-ii_en_c_07121990_redemptoris-missio.html.

⁴ Tamże, nr 37.

Jego objawienie i nawiązał z Nim dialog miłości. Z powodu grzechu ta zdolność do dialogu, zarówno na poziomie osobowym, jak społecznym, uległa osłabieniu, przez co ludzie zaznali i nadal zaznają goryczy niezrozumienia i wyobcowania. Bóg ich jednak nie opuścił, ale posłał im własnego Syna (por. Mk 12,1-11). W Słowie Wcielonym zjawisko przekazu zyskuje najwyższe znaczenie zbawcze: to dzięki niemu bowiem człowiek otrzymuje w Duchu Świętym zdolność przyjęcia zbawienia, obwieszczenia go braciom i dawania o nim świadectwa.

Komunikacja między Bogiem a ludzkością osiągnęła zatem doskonałość w Słowie Wcielonym. Akt miłości, poprzez który Bóg się objawia, w połączeniu z odpowiedzią wiary ze strony ludzkości, prowadzi do owocnego dialogu. Właśnie dlatego możemy prosić Pana – posługując się słowami prośby uczniów: „Naucz nas modlić się” (Łk 11,1) – ażeby pomagał nam zrozumieć, jak mamy porozumiewać się z Bogiem oraz z ludźmi za pomocą wspaniałych narzędzi przekazu. W perspektywie tej ostatecznej i decydującej komunikacji media okazują się opatrnościowym narzędziem, pozwalającym docierać do ludzi na każdej szerokości geograficznej, przekraczać bariery czasu, przestrzeni i języka, przekazywać na wiele różnych sposobów treści wiary i wskazywać każdemu poszukującemu niezawodną przystań, gdzie można nawiązać dialog z tajemnicą Boga objawioną w pełni w Chrystusie Jezusie.

Słowo Wcielone dało nam przykład, jak mamy porozumiewać się z Ojcem i z ludźmi, zarówno w milczeniu i skupieniu, jak i przez przepowiadanie na każdym miejscu i najróżniejszymi językami. Chrystus wyjaśnia Pisma, wypowiada się w przypowieściach, rozmawia w domowym zaciszu, naucza na placach, na ulicach, na brzegu jeziora i na szczycie góry. Osobiste spotkanie z Nim nie pozostawia obojętnym, ale przeciwnie – przynagla do naśladowania Go: „Co mówię wam w ciemności, powtarzajcie w świetle, a co słyszycie na ucho, rozgłaszajcie na dachach” (Mt 10,27).

Następuje później moment kulminacyjny, kiedy komunikacja staje się pełną komunią w spotkaniu eucharystycznym. Gdy wierzący rozpoznają Jezusa po „łamaniu chleba” (por. Łk 24,30-31), czują się zobowiązani głosić Jego śmierć i zmartwychwstanie, stając się odważnymi i radosnymi świadkami Jego Królestwa (por. Łk 24,35).

Dzięki odkupieniu zdolność wierzących do komunikowania się została uzdrowiona i odnowiona. Spotkanie z Chrystusem czyni z nich nowe stworzenia, pozwala im stać się częścią tego ludu, który On pozyskał dla siebie za cenę własnej krwi, umierając na krzyżu, i wprowadza ich w głębię życia Trójcy, które jest nieustannym i wzajemnym przekazywaniem doskonałej i nieskończonej miłości między Ojcem, Synem i Duchem Świętym.

Komunikacja przenika istotne wymiary życia Kościoła, powołanego, aby głosić wszystkim radosne orędzie zbawienia. Dlatego korzysta on z możliwo-

ści, jakie stwarzają środki społecznego przekazu, widząc w nich opatrnościowy dar Boży dla naszej epoki, który ma wzmocnić komunie i pozwolić na bardziej skuteczne przepowiadanie⁵. Media pozwalają ukazać powszechny charakter Ludu Bożego, sprzyjając częstszej i bardziej bezpośredniej łączności między Kościołami lokalnymi, wspomagając wzajemne poznanie i współpracę.

Dziękujemy Bogu za istnienie tych niezwykle skutecznych narzędzi, które w rękach ludzi wierzących, obdarzonych geniuszem wiary i kierujących się światłem Ducha Świętego, mogą ułatwiać szerzenie Ewangelii i zacieśniać więzi jedności między wspólnotami kościelnymi.

Środki przekazu są dla Kościoła cenną pomocą w szerzeniu Ewangelii i wartości religijnych, w rozwijaniu dialogu oraz współpracy ekumenicznej i międzyreligijnej, a także w obronie trwałych zasad, które są niezbędne przy budowie społeczeństwa szanującego godność człowieka i dbającego o wspólne dobro. Kościół chętnie się nimi posługuje, aby dostarczać informacji o sobie i poszerzać zasięg ewangelizacji, katechezy i formacji. Korzystanie z nich traktuje jako odpowiedź na polecenie Chrystusa: „Idźcie na cały świat i głoście Ewangelię wszelkiemu stworzeniu!” (Mk 16,15).

Jest to z pewnością misja niełatwa w naszej epoce, w której coraz powszechniejsze staje się przekonanie, że czas pewników należy nieodwołalnie do przeszłości; wielu jest zdania, że człowiek powinien nauczyć się żyć ze świadomością całkowitego braku sensu, pod znakiem tymczasowości i ulotności⁶. W takim kontekście środki społecznego przekazu mogą być wykorzystywane zarówno „aby głosić Ewangelię, jak i żeby zmusić ją do milczenia w ludzkich sercach”⁷. Jest to poważne wyzwanie dla wierzących, zwłaszcza dla rodziców, rodzin i dla wszystkich odpowiedzialnych za wychowanie dzieci i młodzieży. Z duszpasterską rozważą i mądrością należy zachęcać członków kościelnej wspólnoty, którzy odznaczają się szczególnymi uzdolnieniami do pracy w dziedzinie przekazu, aby zdobywali kwalifikacje zawodowe, pozwalające prowadzić dialog z rozległym światem mediów.

Docenienie wartości mediów nie jest jednak wyłącznie zadaniem „specjalistów” w tej dziedzinie, ale całej wspólnoty kościelnej. Jeśli – jak już powiedzieliśmy – środki społecznego przekazu służą różnym formom wyrażania wiary, chrześcijanie muszą brać pod uwagę oddziaływanie kultury medialnej,

⁵ Por. Jan Paweł II, Adhortacja apostolska *Christifideles laici*, nr 18-24, w: *Adhortacje Ojca Świętego Jana Pawła II*, Wydawnictwo Znak, Kraków 1996, s. 376-390; zob. Papieska Rada ds. Środków Społecznego Przekazu, Instrukcja duszpasterska *Aetatis novae* o przekazie społecznym w dwudziestą rocznicę ogłoszenia instrukcji *Communio et progressio*, „L'Osservatore Romano” 13(1992) nr 6(143), s. 50-59.

⁶ Por. Jan Paweł II, Encyklika *Fides et ratio*, nr 91, The Holy See, https://www.vatican.va/content/john-paul-ii/pl/encyclicals/documents/hf_jp-ii_enc_14091998_fides-et-ratio.html.

⁷ Por. *Aetatis novae*, nr 4, s. 52.

w której żyją: od liturgii, która jest najwyższym i podstawowym wyrazem łączności z Bogiem i z braćmi, po katechezę, która nie może ignorować faktu, że zwraca się do ludzi będących pod wpływem współczesnej kultury i jej języków.

Obecne zjawisko, jakim jest rozwój przekazu społecznego, skłania Kościół do przeprowadzenia swoistej „rewizji” duszpasterskiej i kulturowej, by w odpowiedni sposób stawić czoło epokowym przemianom, których jesteśmy świadkami. Zadanie to powinni podjąć przede wszystkim pasterze: należy w istocie zabiegać o to, aby głoszenie Ewangelii dokonywało się w sposób jasny, który będzie zachęcał do jej słuchania i sprzyjał jej przyjęciu⁸. Szczególna odpowiedzialność w tej dziedzinie spoczywa na osobach konsekrowanych – ze zgromadzeń, których charyzmat wiąże się z działalnością w sferze przekazu społecznego. Ukształtowane duchowo i zawodowo, „niech włączają się aktywnie – w miarę potrzeb duszpasterskich – [...] aby z jednej strony zapobiegać szkodom powodowanym przez niewłaściwe ich używanie, a z drugiej przyczyniać się do podnoszenia poziomu programów przez treści zgodne z prawem moralnym i bogate w wartości humanistyczne i chrześcijańskie”⁹.

Właśnie ze względu na doniosłe znaczenie mediów już piętnaście lat temu uznałem za niewskazane pozostawianie tej dziedziny inicjatywie pojedynczych osób lub niewielkich zespołów i zalecałem, aby przyznano im właściwe miejsce w programach duszpasterskich¹⁰. Zwłaszcza nowe techniki stwarzają dodatkowe możliwości komunikacji, pojmowanej jako służba wspomagająca pasterskie rządy oraz organizację rozlicznych zadań wspólnoty chrześcijańskiej. Przytoczmy jako przykład Internet, który nie tylko stwarza większe możliwości zdobywania informacji, ale uczy także przekazu interaktywnego¹¹. Wielu chrześcijan używa już twórczo tego nowego narzędzia, odkrywając jego przydatność w ewangelizacji, kształceniu, komunikacji wewnętrznej, administracji i zarządzaniu. Oprócz Internetu należy też korzystać z innych nowych środków, jak również próbować wszystkich możliwych zastosowań tradycyjnych metod przekazu. Dzienniki i czasopisma, różnego rodzaju publikacje, katolickie stacje radiowe i telewizyjne są nadal niezwykle przydatne w całym systemie kościelnego przekazu.

Treści przekazu powinny być oczywiście dostosowane do potrzeb poszczególnych środowisk, natomiast jego celem powinno być zawsze uświadamianie

⁸ Por. J a n P a w e ł II, Adhortacja apostolska *Pastores gregis*, nr 30, „L'Osservatore Romano” wyd. pol. 25(2004) nr 1(259), s. 22.

⁹ T e n ż e, Adhortacja apostolska *Vita consecrata*, nr 99, w: *Adhortacje Ojca Świętego Jana Pawła II*, s. 931n.

¹⁰ Por. t e n ż e, *Redemptoris missio*, nr 37.

¹¹ Por. Papieska Rada ds. Środków Społecznego Przekazu, *Kościół a Internet*, nr 6, Libreria Editrice Vaticana, Città del Vaticano 2002, s. 13-15.

ludziom etycznego i moralnego wymiaru informacji¹². Podobnie też ważne jest zapewnienie formacji i opieki duszpasterskiej osobom pracującym zawodowo w dziedzinie przekazu społecznego. Często spotykają się one w codziennej pracy z silnymi naciskami i z dylematami etycznymi; wiele z nich „szczerze pragnie wiedzieć, co jest etycznie i moralnie godziwe, i zgodnie z tym postępować”, i oczekuje od Kościoła wskazań i pomocy¹³.

Kościół, który na mocy orędzia zbawienia powierzonego mu przez Chrystusa jest także nauczycielem ludzkości, uważa za swój obowiązek przyczynianie się do coraz lepszego rozumienia, jakie możliwości i odpowiedzialność wiążą się z obecnym rozwojem środków społecznego przekazu. Właśnie dlatego, że środki przekazu oddziałują na sumienie jednostek, kształtują ich mentalność i wpływają na ich wizję rzeczywistości, trzeba jasno i zdecydowanie przypominać, że są one dziedzictwem, które należy chronić i rozwijać. Konieczne jest, aby także sfera przekazu społecznego została włączona w organiczną strukturę praw i obowiązków, zarówno w tym, co dotyczy formacji i odpowiedzialności etycznej, jak i w odniesieniu do praw i działania jako instytucji.

Za właściwy rozwój środków przekazu w służbie wspólnego dobra odpowiedzialni są wszyscy i każdy osobiście¹⁴. Ze względu na ścisłe powiązanie mediów ze sferą gospodarki, polityki i kultury potrzebny jest system zarządzania nimi, zdolny zagwarantować centralne miejsce i godność człowieka, pierwszeństwo rodziny – podstawowej komórki społecznej, oraz poprawne relacje między różnymi podmiotami.

Konieczne jest dokonanie pewnych wyborów w trzech podstawowych dziedzinach: formacja, współuczestnictwo, dialog.

Przede wszystkim potrzebna jest szeroko zakrojona działalność formacyjna, szerząca właściwą wiedzę o mediach i ucząca świadomego korzystania z nich. Wprowadzane przez media nowe języki przekazu modyfikują procesy uczenia się i zmieniają jakość relacji międzyludzkich, toteż brak należytej formacji może stworzyć niebezpieczeństwo, że zamiast służyć ludziom, będą one posługiwać się nimi i wywierać na nich przemożny wpływ. Dotyczy to zwłaszcza młodych, którzy w sposób naturalny okazują zainteresowanie nowościami technicznymi i właśnie dlatego tym bardziej potrzebują przygotowania do odpowiedzialnego i krytycznego korzystania z mediów.

Następnie chciałbym zwrócić uwagę na sprawę dostępu do mediów i odpowiedzialnego współudziału w zarządzaniu nimi. Jeżeli środki społecznego przekazu są dobrem przeznaczonym dla całej ludzkości, należy szukać coraz

¹² Por. *Inter mirifica*, nr 15-16, s. 92n.; Papieska Komisja Środków Przekazu Społecznego, Instrukcja duszpasterska *Communio et progressio*, nr 107, w: „Acta Apostolicae Sedis” 63(1971), s. 631; *Aetatis novae*, nr 18, s. 56.

¹³ *Aetatis novae*, nr 19, s. 56.

¹⁴ Zob. *Katechizm Kościoła katolickiego*, nr 2494, Pallottinum, Poznań 1994, s. 559.

bardziej aktualnych form kierowania nimi, umożliwiających szeroki udział społeczny, także poprzez stosowne działania prawodawcze. Trzeba zabiegać o rozwój kultury współodpowiedzialności.

Na koniec, nie można zapominać o wielkim potencjale mediów w zakresie umożliwiania dialogu, dzięki czemu media stają się środkami służącymi poznaniu, solidarności i pokojowi. Media stanowią ogromne bogactwo, jeśli służą porozumieniu między narodami; ale stają się niszczącą „bronią”, jeśli służą utrzymywaniu sytuacji niesprawiedliwości i podsycaniu konfliktów. Już mój czcigodny poprzednik bł. Jan XXIII w Encyklice *Pacem in terris*¹⁵ proroczo przestrzegł ludzkość przed tymi zagrożeniami¹⁶.

Wielkie zainteresowanie budzi refleksja nad rolą „opinii publicznej w Kościele” i „Kościoła w opinii publicznej”. Podczas spotkania z wydawcami katolickich czasopism mój czcigodny poprzednik Pius XII powiedział, że w życiu Kościoła brakowałoby czegoś, gdyby nie było opinii publicznej. Ta sama idea została sformułowana także w innych okolicznościach¹⁷, zaś Kodeks Prawa Kanonicznego uznaje – pod pewnymi warunkami – prawo do wyrażania własnej opinii¹⁸. Choć jest prawdą, że prawdy wiary nie podlegają arbitralnym interpretacjom, a poszanowanie praw innych ludzi wyznacza naturalne granice wyrażaniu własnych ocen, prawdą jest również, że w innych dziedzinach istnieje przestrzeń wymiany opinii między katolikami w formie dialogu respektującego nakazy sprawiedliwości i roztropności.

Komunikacja, zarówno wewnątrz kościelnej wspólnoty, jak i Kościoła ze światem, musi odznaczać się przejrzystością i wymaga nowego podejścia do kwestii związanych ze środkami przekazu. Komunikacja ta winna prowadzić do konstruktywnego dialogu, który będzie sprzyjał kształtowaniu się w chrześcijańskiej wspólnotie opinii publicznej opartej na rzetelnej informacji i zdol-

¹⁵ Zob. Jan XXIII, Encyklika *Pacem in terris*, w: *Dokumenty nauki społecznej Kościoła*, red. M. Radwan, L. Dyczewski, A. Stanowski, Fundacja Jana Pawła II – Redakcja Wydawnictw KUL, t. 1, s. 269-312.

¹⁶ Zob. Jan Paweł II, *Środki społecznego przekazu w służbie prawdziwego pokoju w świetle encykliki „Pacem in terris”*. Orędzie na XXXVII Światowy Dzień Środków Społecznego Przekazu (Watykan, 24 I 2003), „L'Osservatore Romano” wyd. pol. 24(2003) nr 3(251), s. 7-9.

¹⁷ Por. Sobór Watykański II, Konstytucja dogmatyczna o Kościele *Lumen gentium*, nr 37, w: Sobór Watykański II, *Konstytucje. Dekrety. Deklaracje*, s. 140; *Communio et progressio*, nr 114-117, s. 634n.

¹⁸ „Stosownie do posiadanej wiedzy, kompetencji i znaczenia, przysługuje im prawo, a niekiedy nawet obowiązek, wyjawiania swojego zdania świętym pasterzom w sprawach dotyczących dobra Kościoła oraz – zachowując nienaruszalność wiary i obyczajów, szacunek dla pasterzy, biorąc pod uwagę wspólny pożytek i godność osoby – podawania go do wiadomości innym wiernym”. Kodeks Prawa Kanonicznego kan. 212, § 3, Archidiecezja Lubelska, „Uaktualnione tłumaczenie Kodeksu Prawa Kanonicznego – 14 III 2022”, <https://archidiecezjalubelska.pl/blog/uaktualnione-tlumaczenie-polskie-kodeksu-prawa-kanonicznego-14-03-2022/>). Por. Kodeks Kanonów Kościołów Wschodnich, kan. 15, § 3.

nej do rozeznawania problemów. Kościół odczuwa potrzebę informowania o swojej działalności i ma do tego prawo, podobnie jak inne instytucje i grupy, zarazem jednak musi mieć możliwość zapewnienia sobie – gdy jest to konieczne – należytej dyskrecji, która wszakże nie powinna stanowić przeszkody dla rzetelnego i wystarczającego informowania o wydarzeniach kościelnych. Jest to jedna z dziedzin, w których najbardziej potrzebna jest współpraca między wiernymi świeckimi a pasterzami, jak bowiem słusznie podkreśla Sobór: „Po tym zażyłym obcowaniu ze sobą świeckich i pasterzy należy się spodziewać rozlicznych dóbr dla Kościoła: dzięki temu bowiem w świeckich umacnia się poczucie własnej odpowiedzialności, wzmacnia się zapał i siły ludzi świeckich łatwiej łączą się z pracą pasterzy. Ci zaś, wsparci doświadczeniem świeckich, mogą zdobyć celniejszy i odpowiedniejszy osąd zarówno w sprawach duchowych, jak i doczesnych, tak aby cały Kościół, umocniony przez wszystkich swoich członków, skuteczniej pełnił swe posłannictwo dla życia świata”¹⁹.

Dla wierzących i ludzi dobrej woli wielkim wyzwaniem w naszych czasach jest zabieganie o swobodny i rzetelny przekaz społeczny, który będzie się przyczyniał do umacniania integralnego postępu świata. Od wszystkich oczekuje się zdolności do wnikliwego rozeznania i nieustannej czujności, kształtowania w sobie zdrowego zmysłu krytycyzmu wobec siły przekonywania środków przekazu.

Wierzący w Chrystusa wiedzą, że także w tej dziedzinie mogą liczyć na pomoc Ducha Świętego. Pomoc szczególnie niezbędną, jeśli weźmiemy pod uwagę, że trudności wynikające z samej natury przekazu społecznego mogą być pogłębiane w znacznym stopniu przez ideologie, żądze zysku i władzy, rywalizacje i konflikty między jednostkami i grupami, a także na skutek ludzkich słabości i różnych form zła społecznego. Nowoczesne techniki w imponujący sposób zwiększają tempo, rozmiar i zasięg przekazu społecznego, ale nie przyczyniają się w takiej samej mierze do wsparcia owego delikatnego dialogu między sercami i umysłami, który winien charakteryzować wszelki przekaz służący solidarności i miłości.

W historii zbawienia Chrystus objawił się nam jako „przekaziciel” Ojca: „W tych ostatecznych dniach [Bóg] przemówił do nas przez Syna” (Hbr 1,2). Chrystus, odwieczne Słowo, które stało się ciałem, gdy mówi, okazuje zawsze szacunek tym, którzy Go słuchają, uczy zrozumienia dla ich sytuacji i potrzeb, współczucia dla ich cierpienia i stanowczości w mówieniu im tego, co powinni usłyszeć, bez narzucania niczego, bez kompromisów, kłamstw i manipulacji. Jezus naucza, że przekaz społeczny jest aktem natury moralnej: „Dobry człowiek z dobrego skarbcza wydobywa dobre rzeczy, a zły człowiek ze złego skarbcza wydobywa złe rzeczy. A powiadam wam: Z każdego bezużytecznego

¹⁹ *Lumen gentium*, nr 37, s. 140n.

słowa, które wypowiedzą ludzie, zdadzą sprawę w dzień sądu. Bo na podstawie słów twoich będziesz uniewinniony i na podstawie słów twoich będziesz potępiony» (Mt 12,35-37).

Apostoł Paweł kieruje bardzo jasne przesłanie do wszystkich, którzy uczestniczą w procesie przekazu społecznego – do polityków, do ludzi pracujących w tej dziedzinie i do odbiorców: „Dlatego odrzuciwszy kłamstwo, niech każdy z was mówi prawdę swemu bliźniemu, bo jesteście nawzajem dla siebie członkami. [...] Niech nie wychodzi z waszych ust żadna mowa szkodliwa, lecz tylko budująca, zależnie od potrzeby, by wyświadczała dobro słuchającym» (Ef 4,25.29).

Do pracowników środków przekazu, a zwłaszcza do ludzi wierzących zatrudnionych w tej ważnej dziedzinie życia społecznego kieruję wezwanie, z jakim na samym początku mojej posługi Pasterza Kościoła powszechnego zwróciłem się do całego świata: „Nie lękajcie się!”

Nie lękajcie się nowych technologii! Są one „wśród niezwyklej wynalazków” – inter mirifica – które Bóg dał nam do dyspozycji, abyśmy mogli odkrywać prawdę, korzystać z niej i ją przekazywać – również prawdę o naszej godności i o przeznaczeniu jako Jego dzieci, jako dziedziców Jego wiekuistego Królestwa.

Nie lękajcie się sprzeciwu świata! Jezus zapewnił nas: „Jam zwyciężył świat!” (J 16,33).

Nie lękajcie się także własnej słabości i niedoskonałości! Boski Mistrz powiedział: „Ja jestem z wami przez wszystkie dni, aż do skończenia świata” (Mt 28,20). Przekazujcie orędzie nadziei, łaski i miłości Chrystusa i podtrzymujcie zawsze w tym przemijającym świecie żywą świadomość wiecznej perspektywy Nieba – perspektywy, której nie dosięgnie nigdy bezpośrednio żaden środek przekazu: „Ani oko nie widziało, ani ucho nie słyszało, ani serce człowieka nie zdołało pojąć, jak wielkie rzeczy przygotował Bóg tym, którzy Go miłują” (1 Kor 2,9).

JOHN PAUL II

THE RAPID DEVELOPMENT*

The rapid development of technology in the area of the media is surely one of the signs of progress in today's society. In view of these innovations in continuous evolution, the words found in the Decree of the Second Vatican Ecumenical Council, *Inter Mirifica*, promulgated by my venerable predecessor, the servant of God Paul VI, December 4, 1963, appear even more pertinent: "Man's genius has with God's help produced marvelous technical inventions from creation, especially in our times. The Church, our mother, is particularly interested in those which directly touch man's spirit and which have opened up new avenues of easy communication of all kinds of news, of ideas and orientations."¹

More than forty years after the publication of that document, it appears appropriate to reflect on the "challenges" which the communications media constitute for the Church, which Paul VI said "would feel guilty before the Lord if she did not utilize these powerful means."² In fact, the Church is not only called upon to use the mass media to spread the Gospel but, today more than ever, to integrate the message of salvation into the "new culture" that these powerful means of communication create and amplify. It tells us that the use

* Extracts from John Paul II's Apostolic Letter *The Rapid Development* to Those Responsible for Communications (Vatican, 24 January 2005).

For the original text see "Apostolic Letter *The Rapid Development* of the Holy Father John Paul II to Those Responsible for Communications," The Holy See, https://www.vatican.va/content/john-paul-ii/en/apost_letters/2005/documents/hf_jp-ii_apl_20050124_il-rapido-sviluppo.html.

© Copyright Dicastero per la Comunicazione and Libreria Editrice Vaticana.

¹ The Second Vatican Council, "Decree on the Media of Social Communications *Inter Mirifica*," Section 1, The Holy See https://www.vatican.va/archive/hist_councils/ii_vatican_council/documents/vat-ii_decree_19631204_inter-mirifica_en.html.

² P a u l VI, "Apostolic Exhortation *Evangelii Nuntiandi*," Section 45, The Holy See, https://www.vatican.va/content/paul-vi/en/apost_exhortations/documents/hf_p-vi_exh_19751208_evangelii-nuntiandi.html.

of the techniques and the technologies of contemporary communications is an integral part of its mission in the third millennium.

Moved by this awareness, the Christian community has taken significant steps in the use of the means of communication for religious information, for evangelization and catechesis, for the formation of pastoral workers in this area, and for the education to a mature responsibility of the users and the recipients of the various communications media.

Many challenges face the new evangelization in a world rich with communicative potential like our own. Because of this, I wanted to underline in the Encyclical *Redemptoris Missio*³ that the first Areopagus of modern times is the *world of communications*, which is capable of unifying humanity and transforming it into—as it is commonly referred to—“a global village.”⁴ The communications media have acquired such importance as to be the principal means of guidance and inspiration for many people in their personal, familial, and social behavior. We are dealing with a complex problem, because the culture itself, prescinding from its content, arises from the very existence of new ways to communicate with hitherto unknown techniques and vocabulary.

Ours is an age of global communication in which countless moments of human existence are either spent with, or at least confronted by, the different processes of the mass media. I limit myself to mentioning the formation of personality and conscience, the interpretation and structuring of affective relationships, the coming together of the educative and formative phases, the elaboration and diffusion of cultural phenomena, and the development of social, political and economic life.

The mass media can and must promote justice and solidarity according to an organic and correct vision of human development, by reporting events accurately and truthfully, analyzing situations and problems completely, and providing a forum for different opinions. An authentically ethical approach to using the powerful communication media must be situated within the context of a mature exercise of freedom and responsibility, founded upon the supreme criteria of truth and justice.

The world of mass media also has need of Christ’s redemption. To analyze with the eyes of faith the processes and value of communications, the deeper appreciation of Sacred Scripture can undoubtedly help as a “great code” of communication of a message which is not ephemeral, but fundamental for its saving value.

³ See John Paul II, “Encyclical Letter *Redemptoris Missio* on the Permanent Validity of the Church’s Missionary Mandate,” The Holy See, https://www.vatican.va/content/john-paul-ii/en/encyclicals/documents/hf_jp-ii_enc_07121990_redemptoris-missio.html.

⁴ Ibidem, Section 37.

Salvation History recounts and documents the communication of God with man, a communication which uses all forms and ways of communicating. The human being is created in the image and likeness of God in order to embrace divine revelation and to enter into loving dialogue with Him. Because of sin, this capacity for dialogue at both the personal and social level has been altered, and humanity has had to suffer, and will continue to suffer, the bitter experience of incomprehension and separation. God, however, did not abandon the human race, but sent his own Son (see Mk 12:1–11). In the Word made flesh communication itself takes on its most profound saving meaning: thus, in the Holy Spirit, the human being is given the capacity to receive salvation, and to proclaim and give witness to it before the world.

The communication between God and humanity has thus reached its perfection in the Word made flesh. The act of love by which God reveals himself, united to the response of faith by humanity, generates a fruitful dialogue. Precisely for this reason, making our own in a certain sense the request of the disciples, “teach us to pray” (Lk 11:1), we can ask the Lord to help us to understand how to communicate with God and with other human beings through the marvelous communications media. In light of so decisive and definitive a communication, the media provide a providential opportunity to reach people everywhere, overcoming barriers of time, of space and of language; presenting the content of faith in the most varied ways imaginable; and offering to all who search the possibility of entering into dialogue with the mystery of God, revealed fully in Christ Jesus.

The Incarnate Word has left us an example of how to communicate with the Father and with humanity, whether in moments of silence and recollection, or in preaching in every place and in every way. He explains the Scriptures, expresses himself in parables, dialogues within the intimacy of the home, speaks in the squares, along the streets, on the shores of the lake and on the mountaintops. The personal encounter with him does not leave one indifferent, but stimulates imitation: “What I say to you in the darkness, speak in the light; what you hear whispered, proclaim on the housetops” (Mt 10:27).

There is, however, a culminating moment in which communication becomes full communion: the Eucharistic encounter. By recognizing Jesus in the “breaking of the bread,” (see Lk 24:30–31), believers feel themselves urged on to announce his death and resurrection, and to become joyful and courageous witnesses of his Kingdom (see Lk 24:35).

Thanks to the Redemption, the communicative capacity of believers is healed and renewed. The encounter with Christ makes them new creatures, and permits them to become part of that people which he, dying on the Cross, has won through his blood, and introduces them into the intimate life of the Trinity, which is continuous and circular communication of perfect and infinite love among the Father, the Son and the Holy Spirit.

Communication permeates the essential dimensions of the Church, which is called to announce to all the joyful message of salvation. For this reason, the Church takes advantage of the opportunities offered by the communications media as pathways providentially given by God to intensify communion and to render more penetrating the proclamation of His word.⁵ The media permit the manifestation of the universal character of the People of God, favoring a more intense and immediate exchange among local Churches, and nourishing mutual awareness and cooperation.

We give thanks to God for the presence of these powerful media which, if used by believers with the genius of faith and in docility to the light of the Holy Spirit, can facilitate the communication of the Gospel and render the bonds of communion among ecclesial communities more effective.

In the communications media the Church finds a precious aid for spreading the Gospel and religious values, for promoting dialogue, ecumenical and inter-religious cooperation, and also for defending those solid principles which are indispensable for building a society which respects the dignity of the human person and is attentive to the common good. The Church willingly employs these media to furnish information about itself and to expand the boundaries of evangelization, of catechesis and of formation, considering their use as a response to the command of the Lord: "Go into the whole world and proclaim the gospel to every creature" (Mk 16:15).

This is certainly not an easy mission in an age such as ours, in which there exists the conviction that the time of certainties is irretrievably past. Many people, in fact, believe that humanity must learn to live in a climate governed by an absence of meaning, by the provisional and by the fleeting.⁶ In this context, the communications media can be used "to proclaim the Gospel or to reduce it to silence within men's hearts."⁷ This poses a serious challenge for believers, especially for parents, families and all those responsible for the formation of children and young people. Those individuals in the Church community particularly gifted with talent to work in the media should be encouraged with

⁵ See John Paul II, "Apostolic Exhortation *Christifideles Laici* on the Vocation and the Mission of the Lay Faithful in the Church and in the World," Sections 18–24, The Holy See, https://www.vatican.va/content/john-paul-ii/en/apost_exhortations/documents/hf_jp-ii_exh_30121988_christifideles-laici.html. See also Pontifical Council for Social Communications, "Pastoral Instruction *Aetatis Novae* on Social Communications on the Twentieth Anniversary of *Communio et Progressio*," Section 10, The Holy See, https://www.vatican.va/roman_curia/pontifical_councils/pccs/documents/rc_pc_pccs_doc_22021992_aetatis_en.html.

⁶ See John Paul II, "Encyclical Letter *Fides et Ratio* on the Relationship between Faith and Reason," Sections 76–77, The Holy See, https://www.vatican.va/content/john-paul-ii/en/encyclicals/documents/hf_jp-ii_enc_14091998_fides-et-ratio.html.

⁷ *Aetatis Novae*, Section 4.

pastoral prudence and wisdom, so that they may become professionals capable of dialoguing with the vast world of the mass media.

The appreciation of the media is not reserved only to those already adept in the field, but to the entire Church Community. If, as has already been noted, the communications media take into account different aspects of the expression of faith, Christians must take into account the media culture in which they live: from the Liturgy, the fullest and fundamental expression of communication with God and with one another, to Catechesis, which cannot prescind from the fact of being directed to people immersed in the language and the culture of the day.

The current phenomenon of communications impels the Church towards a sort of pastoral and cultural revision, so as to deal adequately with the times in which we live. Pastors, above all, must assume this responsibility. Everything possible must be done so that the Gospel might permeate society, stimulating people to listen to and embrace its message.⁸ Consecrated persons belonging to institutions having the charism of using the mass media have a particular responsibility in this regard. Spiritually and professionally formed towards this end, these institutions, “should willingly lend their help, wherever pastorally appropriate ... in order to offset the inappropriate use of the media and to promote higher quality programs, the contents of which will be respectful of the moral law and rich in human and Christian values.”⁹

Such is the importance of the mass media that fifteen years ago I considered it inopportune to leave their use completely up to the initiatives of individuals or small groups, and suggested that they be decisively inserted into pastoral programs.¹⁰ New technologies, in particular, create further opportunities for communication understood as a service to the pastoral government and organization of the different tasks of the Christian community. One clear example today is how the Internet not only provides resources for more information, but habituates persons to interactive communication.¹¹ Many Christians are already creatively using this instrument, exploring its potential to assist in the tasks of evangelization and education, as well as of internal communication,

⁸ See J o h n P a u l I I, “Post-Synodal Apostolic Exhortation, *Pastores Gregis* on the Bishop, Servant of the Gospel of Jesus Christ for the Hope of the World,” Section 30, The Holy See, https://www.vatican.va/content/john-paul-ii/en/apost_exhortations/documents/hf_jp-ii_exh_20031016_pastores-gregis.html.

⁹ See J o h n P a u l I I, “Apostolic Exhortation *Vita Consecrata* on the Consecrated Life and Its Mission in the Church and in the World,” Section 99, The Holy See, https://www.vatican.va/content/john-paul-ii/en/apost_exhortations/documents/hf_jp-ii_exh_25031996_vita-consecrata.html.

¹⁰ J o h n P a u l I I, *Redemptoris Missio*, Section 37.

¹¹ See Pontifical Council for Social Communications, *The Church and Internet*, Section 6, The Holy See, https://www.vatican.va/roman_curia/pontifical_councils/pccs/documents/rc_pc_pccs_doc_20020228_church-internet_en.html.

administration and governance. However, alongside the Internet, other new means of communication, as well as traditional ones, should be used. Daily and weekly newspapers, publications of all types, and Catholic television and radio still remain highly useful means within a complete panorama of Church communications.

While the content being communicated must obviously be adapted to the needs of different groups, the goal must always be to make people aware of the ethical and moral dimension of the information.¹² In the same way, it is important to assure that media professionals receive the necessary formation and pastoral attention to confront the particular tensions and ethical dilemmas that arise in their daily work. Often these men and women “sincerely desire to know and practice what is ethically and morally just,”¹³ and look to the Church for guidance and support.

The Church, which in light of the message of salvation entrusted to it by the Lord is also a teacher of humanity, recognizes the duty to offer its own contribution for a better understanding of outlooks and responsibilities connected with current developments in communications. Especially because these influence the consciences of individuals, form their mentality and determine their view of things, it is important to stress in a forceful and clear way that the mass media constitute a patrimony to safeguard and promote. The communications media must enter into the framework of organically structured rights and duties, be it from the point of view of formation and ethical responsibility, or from reference to laws and institutional codes.

The positive development of the media at the service of the common good is a responsibility of each and every one.¹⁴ Because of the close connections the media have with economics, politics and culture, there is required a management system capable of safeguarding the centrality and dignity of the person, the primacy of the family as the basic unit of society and the proper relationship among them.

We are faced with three fundamental options: *formation, participation, and dialogue*.

In the first place, a *vast work of formation* is needed to assure that the mass media be known and used intelligently and appropriately. The new vocabulary they introduce into society modifies both learning processes and the quality of

¹² See *Inter Mirifica*, Sections 15–16; Pontifical Council for Social Communications, “Pastoral Instruction *Communio et Progressio* on the Means of Social Communication,” Section 107, The Holy See, https://www.vatican.va/roman_curia/pontifical_councils/pccs/documents/rc_pc_pccs_doc_23051971_communio_en.html; *Aetatis Novae*, Section 18.

¹³ *Aetatis Novae*, Section 19.

¹⁴ See *The Catechism of the Catholic Church*, Section 2494, The Holy See, https://www.vatican.va/archive/ENG0015/_INDEX.HTM.

human relations, so that, without proper formation, these media run the risk of manipulating and heavily conditioning, rather than serving people. This is especially true for young people, who show a natural propensity towards technological innovations, and as such are in even greater need of education in the responsible and critical use of the media.

In the second place, I would like to recall our attention to the subject of media access, and of *co-responsible participation* in their administration. If the communications media are a good destined for all humanity, then ever-new means must be found—including recourse to opportune legislative measures—to make possible a true participation in their management by all. The culture of co-responsibility must be nurtured.

Finally, there cannot be forgotten the great possibilities of mass media in promoting dialogue, becoming vehicles for reciprocal knowledge, of solidarity and of peace. They become a powerful resource for good if used to foster understanding between peoples; a destructive “weapon” if used to foster injustice and conflicts. My venerable predecessor, Blessed John XXIII, already prophetically warned humanity of such potential risks in the Encyclical, *Pacem in Terris*.¹⁵

The reflection upon the role “of public opinion in the Church,” and “of the Church in public opinion” aroused great interest. In a meeting with the editors of Catholic publications, my venerable predecessor, Pius XII, stated that something would be missing from the life of the Church were it not for public opinion. This same idea has since been repeated on other occasions¹⁶, and in the Code of Canon Law there is recognized, under certain conditions, the right to the expression of one’s own opinion.¹⁷ While it is true that the truths of the faith are not open to arbitrary interpretations, and that respect for the rights of others places intrinsic limits upon the expression of one’s judgments, it is no

¹⁵ See J o h n XXIII, „Encyclical Letter *Pacem in Terris* on Establishing Universal Peace in Truth, Justice, Charity, and Liberty,” The Holy See, https://www.vatican.va/content/john-xxiii/en/encyclicals/documents/hf_j-xxiii_enc_11041963_pacem.html. See also J o h n P a u l II, “Message for the 37th World Communications Day, 1 June 2003, ‘The Communications Media at the Service of Authentic Peace in the Light of *Pacem in Terris*,’” The Holy See, https://www.vatican.va/content/john-paul-ii/en/messages/communications/documents/hf_jp-ii_mes_20030124_world-communications-day.html.

¹⁶ See The Second Vatican Council, “Dogmatic Constitution on the Church *Lumen Gentium*,” Section 37, The Holy See, https://www.vatican.va/archive/hist_councils/ii_vatican_council/documents/vat-ii_const_19641121_lumen-gentium_en.html; *Communio et Progressio*, Sections 114–117.

¹⁷ “According to the knowledge, competence, and prestige which they possess, they have the right and even at times the duty to manifest to the sacred pastors their opinion on matters which pertain to the good of the Church and to make their opinion known to the rest of the Christian faithful, without prejudice to the integrity of faith and morals, with reverence toward their pastors, and attentive to common advantage and the dignity of persons.” *Code of Canon Law*, Can. 212, §3, The Holy See, https://www.vatican.va/archive/cod-iuris-canonici/cic_index_en.html.

less true that there is still room among Catholics for an exchange of opinions in a dialogue which is respectful of justice and prudence.

Communication both within the Church community, and between the Church and the world at large, requires openness and a new approach towards facing questions regarding the world of media. This communication must tend towards a constructive dialogue, so as to promote a correctly-informed and discerning public opinion within the Christian community. The Church, like other institutions and groups, has the need and the right to make its activities known. However, when circumstances require, it must be able to guarantee an adequate confidentiality, without thereby prejudicing a timely and sufficient communication about Church events. This is one of the areas in which collaboration between the lay faithful and Pastors is most needed, as the Council appropriately emphasized, "A great many wonderful things are to be hoped for from this familiar dialogue between the laity and their spiritual leaders: in the laity a strengthened sense of personal responsibility; a renewed enthusiasm; a more ready application of their talents to the projects of their spiritual leaders. The latter, on the other hand, aided by the experience of the laity, can more clearly and more incisively come to decisions regarding both spiritual and temporal matters. In this way, the whole Church, strengthened by each one of its members, may more effectively fulfill its mission for the life of the world."¹⁸

The great challenge of our time for believers and for all people of good will is that of maintaining truthful and free communication which will help consolidate integral progress in the world. Everyone should know how to foster an attentive discernment and constant vigilance, developing a healthy critical capacity regarding the persuasive force of the communications media.

Also in this field, believers in Christ know that they can count upon the help of the Holy Spirit. Such help is all the more necessary when one considers how greatly the obstacles intrinsic to communication can be increased by ideologies, by the desire for profit or for power, and by rivalries and conflicts between individuals and groups, and also because of human weakness and social troubles. The modern technologies increase to a remarkable extent the speed, quantity and accessibility of communication, but they above all do not favor that delicate exchange which takes place between mind and mind, between heart and heart, and which should characterize any communication at the service of solidarity and love.

Throughout the history of salvation, Christ presents himself to us as the "communicator" of the Father: "God, in these last days, has spoken to us through his Son" (Heb 1:2). The eternal Word made flesh, in communicating

¹⁸ *Lumen Gentium*, Section 37.

Himself, always shows respect for those who listen, teaches understanding of their situation and needs, is moved to compassion for their suffering and to a resolute determination to say to them only what they need to hear without imposition or compromise, deceit or manipulation. Jesus teaches that communication is a moral act, “A good person brings forth good out of a store of goodness, but an evil person brings forth evil out of a store of evil. I tell you, on the Day of Judgment people will render an account for every careless word they speak. By your words you will be acquitted, and by your words you will be condemned” (Mt 12:35–37).

The apostle Paul has a clear message for those engaged in communications (politicians, professional communicators, spectators), “Therefore, putting away falsehood, speak the truth, each one to his neighbor, for we are members one of another.... No foul language should come out of your mouths, but only such as is good for needed edification, that it may impart grace to those who hear” (Eph 4:25,29).

To those working in communication, especially to believers involved in this important field of society, I extend the invitation which, from the beginning of my ministry as Pastor of the Universal Church, I have wished to express to the entire world “Do not be afraid!”

Do not be afraid of new technologies! These rank “among the marvelous things”—*inter mirifica*—which God has placed at our disposal to discover, to use and to make known the truth, also the truth about our dignity and about our destiny as his children, heirs of his eternal Kingdom.

Do not be afraid of being opposed by the world! Jesus has assured us, “I have conquered the world!” (Jn 16:33)

Do not be afraid even of your own weakness and inadequacy! The Divine Master has said, “I am with you always, until the end of the world” (Mt 28:20). Communicate the message of Christ’s hope, grace and love, keeping always alive, in this passing world, the eternal perspective of heaven, a perspective which no communications medium can ever directly communicate, “What eye has not seen, and ear has not heard, and what has not entered the human heart, what God has prepared for those who love him” (1 Cor 2:9).

KU ZROZUMIENIU SZTUCZNEJ INTELIGENCJI

Zbigniew ORBIK

W POSZUKIWANIU FILOZOFICZNYCH ŹRÓDEŁ PROBLEMATYKI SZTUCZNEJ INTELIGENCJI

Rozwój sztucznej inteligencji implikuje problemy natury filozoficznej. Wielu filozofów od czasów starożytnych stawiało pytania istotne z punktu widzenia metodologii sztucznej inteligencji. Dotyczyły one między innymi natury procesów poznawczych, możliwości konstrukcji języków formalnych opisujących w ścisły sposób rzeczywistość czy idei automatyzacji rozumowań. Refleksja filozoficzna wpisana jest niejako w naturę zagadnień pojawiających się w trakcie rozwoju badań nad sztuczną inteligencją.

Pojawienie się sztucznej inteligencji stanowi jeden z najważniejszych momentów w dotychczasowym rozwoju naukowo-technologicznym. Sam termin „sztuczna inteligencja” – mimo że używany nader często w rozmaitych kontekstach – nie doczekał się dotychczas powszechnie akceptowanej definicji. Wynika to głównie z powodu trudności ze sprecyzowaniem pojęcia samej inteligencji. W definicjach podkreśla się najczęściej, że jest to pewna zdolność do przetwarzania informacji na poziomie abstrakcyjnym, adaptacji do nowych warunków otoczenia, uczenia się czy rozwiązywania skomplikowanych zagadnień natury teoretycznej lub praktycznej. Pojęcie inteligencji implikuje zdolność myślenia. We współczesnych ujęciach określenie to rozszerzono poza sferę intelektualną również na sferę emocjonalną. Tak rozumianą inteligencję przypisać można nie tylko człowiekowi, ale także niektórym wyżej rozwiniętym zwierzętom.

Termin „sztuczna inteligencja” (ang. artificial intelligence – AI) po raz pierwszy pojawił się w roku 1955 w projekcie badawczym profesora Dartmouth College Johna McCarthy’ego¹, poświęconym możliwości stworzenia maszyny symulującej ludzką inteligencję. To pojęcie odnosi się do dziedziny informatyki, której celem jest stworzenie inteligentnych, zatem także myślących maszyn. W tradycji filozoficznej inteligencja (intelekt) jest zazwyczaj przeciwstawiana intuicji. Pojęcie sztucznej inteligencji jest w pełni uprawnione o tyle, o ile obdarzonej nią maszynie (komputerowi) przypisać można funkcje myślenia charakteryzujące umysł ludzki – funkcje mentalne. W ten sposób pytanie o istotę sztucznej inteligencji prowadzi do pytania o istotę świadomości.

¹ Por. J. M c C a r t h y, i in., *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, August 31, 1955, „AI Magazine” 27(2006) nr 4, s. 12.

W prezentowanym artykule podejmuje się zagadnienie sztucznej inteligencji w aspekcie filozoficznym. Pomimo że jej powstanie datuje się na drugą połowę dwudziestego stulecia, pewne intuicje, które można określić jako „informatyczną preświadomość”², istniały już wcześniej. Przejawiały się one w postaci idei filozoficznych dotyczących możliwości ujęcia rozumowań w postaci symbolicznej, natury uzyskiwanej wiedzy, jej zapisu i przechowywania. Były to, innymi słowy, idee, których realizacja praktyczna stała się możliwa dopiero po osiągnięciu pewnego poziomu rozwoju technologicznego. Warto jednak pamiętać, że postęp ten poprzedzony był wielowiekową historią zmagania się umysłu ludzkiego z najbardziej fundamentalnymi problemami dotyczącymi natury świata i jego poznania oraz możliwościami usprawnienia własnych władz poznawczych. Artykuł ma na celu nakreślenie tej drogi oraz wskazanie na wybrane problemy natury filozoficznej, jakie niesie ze sobą rozwój sztucznej inteligencji.

ZARYS HISTORII SZTUCZNEJ INTELIGENCJI

Pojęcie sztucznej inteligencji pojawiło się po raz pierwszy w połowie dwudziestego wieku. W tamtym okresie ukuto używany do dzisiaj termin, lecz sama idea AI jest o wiele starsza. Teoretycznych początków długiego procesu kształtowania się tej idei, który doprowadził do jej urzeczywistnienia, można poszukiwać już w dociekaniach greckich filozofów nad istotą rozumowania, co zwieńczyły osiągnięcia w dziedzinie logiki i matematyki (między innymi pitagorejczyków, Arystotelesa czy Euklidesa). Grecy myśliciele zapoczątkowali przede wszystkim racjonalne myślenie o świecie, czego wynikiem w późniejszych stuleciach było powstanie nowożytnej nauki i techniki, której jednym z rezultatów są obecnie prowadzone badania nad sztuczną inteligencją. Ich rozważania nad strukturą formalną systemów abstrakcyjnych czynią z nich także twórców idei inicjujących proces, który doprowadził po wielu wiekach rozwoju nauki do badań nad sztuczną inteligencją. Jako pierwsi byli bowiem wyrazicielami poglądu, że każde rozumowanie może zostać ujęte w ramach pewnego rachunku (algorytmu).

Badania nad automatyzacją rozumowań i jej skutkami na polu filozofii prowadził w średniowieczu, żyjący na przełomie trzynastego i czternastego wieku, kataloński filozof, teolog i poeta Rajmundus Lullus³. Trudno tutaj

² Por. P. Staciewicz, *Światopogląd informatyczny. Naukowe podstawy i filozoficzne perspektywy*, w: *Informatyka a filozofia. Od informatyki i jej zastosowań do światopoglądu informatycznego*, red. P. Staciewicz, Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa 2015, s. 13.

³ Por. K. Trzęsicki, *Idea sztucznej inteligencji*, „Filozofia i Nauka” 8 (2020) cz. 1, s. 69-96.

oczywiście mówić o badaniach w znaczeniu ściśle technicznym. Lullus był raczej typem wizjonera-marzyciela, dla którego skonstruowana przez niego maszyna logiczna, zwana „volvelle” (łac. *volvere*, pol. obracać), miała być narzędziem rozstrzygania dysput teologicznych. Volvelle może być traktowana jako fizyczna egzemplifikacja idei logicznych Arystotelesa. Konstrukcja tego urządzenia była stosunkowo prosta. Było ono wykonane z papieru lub pergaminu. Rozstrzygając jakiś problem, odwoływano się do zestawu liczb i odpowiadających im symboli. Przymioty Boga reprezentowane były przez dziewięć odpowiadających im liter. Ich kombinacja stanowiła odpowiedź na postawiony problem teologiczny. Były one umieszczone na jednym lub dwóch ruchomych kołach.

Sama idea maszyny Lullusa stanowiła próbę mechanizacji procesu rozumowania. Warto w tym momencie rozróżnić pojęcia mechanizacji i automatyzacji rozumowań. Pierwsze z nich ma swoje źródło w zapoczątkowanej w średniowieczu idei formalizacji rozumowań, która znalazła pełny wyraz we współczesnych systemach logicznych Gottloba Fregego, Giuseppe Peana, Bertranda Russella, Davida Hilberta, Jana Łukasiewicza, Kurta Gödla, Alfreda Tarskiego i innych. W ramach tego nurtu doszło do algebraizacji logiki w postaci rozwiniętej przez George’a Boole’a algebry binarnej, co przyczyniło się do powstania komputerów, które umożliwiły mechanizację rozumowań dedukcyjnych⁴. Mechanizacja rozumowania, ujmując rzecz najprościej, to po prostu techniczna realizacja dowodzenia sformalizowanego, w której używa się nie długopisu czy ołówek, ale impulsów elektrycznych, a adnotacji dokonuje nie ręka człowieka, lecz określony proces fizyczny. Automatyzację rozumowań należy rozumieć jako zastępowanie pracy człowieka w procesie rozumowania przez działanie maszyny bez udziału człowieka; w tym sensie automatyzacja jest kolejnym etapem następującym po mechanizacji. Lullus uczynił zatem ważny krok zmierzający do skonstruowania maszyny, która w założeniu byłaby zdolna do dokonywania pewnych operacji intelektualnych, w szerokim oczywiście tego słowa znaczeniu. Uprawniony wydaje się więc wniosek, że „Lullus jest tym, od którego można rozpocząć historię idei myślących maszyn, a więc historię idei sztucznej inteligencji”⁵. Znaczenie Lullusa polega także na tym, że oddziałał on na wielu późniejszych myślicieli. Zaliczyć do nich można takich uczonych, jak na przykład Giovanni de la Fontana (ok. 1390-1455/56), Mikołaj z Kuzy (1401-1464), Giordano Bruno (1548-1600), Thomas Hobbes (1588-1679) czy Athanasius Kircher (1602-1680). Jak łatwo dostrzec z powyższego zestawienia, były to postaci wielkiej miary, które wywarły znaczący wpływ na europejską filozofię i naukę.

⁴ Por. W. Marciszewski, R. Murawski, *Mechanization of Reasoning in a Historical Perspective*, Editions Rodopi, Amsterdam–Atlanta 1995, s. 12.

⁵ Trzęsicki, dz. cyt., s. 74.

Każdy z wymienionych intelektualistów wniósł wkład w rozwój idei mechanizacji rozumowań i/ ostatecznie powstanie sztucznej inteligencji. Fontana interesował się nie tylko sztuką, ale i inżynierią. Jednym z jego osiągnięć jest opis konstrukcji maszyny ułatwiającej zapamiętywanie oraz wywody dotyczące sztucznej pamięci⁶. Mikołaj z Kuzy, który był autorem licznych dzieł z zakresu matematyki, wzorując się na pomysłach Lullusa, opracował za pomocą ruchomych diagramów i symboli sztukę wnioskowania w warunkach niepewności (łac. *ars coniecturalis*). Interesujące są jego obserwacje dotyczące umysłu, prowadzone w kontekście rozważań na temat różnic zachodzących między umysłem Boga a umysłem człowieka. Mikołaj studiował również możliwość eksperymentalnego zdobywania wiedzy. Można go uznać za prekursora nowoczesnej nauki⁷. Także Bruno nawiązuje do idei *volvelle* Lullusa. W swoich badaniach dążył do stworzenia sztucznej pamięci, którą później wykorzystywał w retoryce. Hobbes formułuje koncepcję operacji syntaktycznych rozumianych jako pewien rodzaj rachunku. Dochodzi on do wniosku, że rozumowanie jest rachunkiem. Gramatyka języka naturalnego to, w jego ujęciu, rachunek dokonywany na słowach. Hobbes pisze: „Gdy człowiek rozumuje, to nie czyni nic innego, tylko ujmuje pojęciowo (*conceive*) całkowitą sumę, dodając składniki, albo też przedstawia sobie pojęciowo (*conceive*) resztę, odejmując jedną sumę od drugiej. Jeżeli jest to wykonywane na słowach, to jest to przedstawieniem sobie pojęciowym (*conceive*) szeregu nazw wszystkich składników i łączeniem ich w nazwę całości; albo też przejściem od nazw całości i jednego składnika do nazwy innego składnika”⁸.

Mało znany współczesnym naukowcom Kircher był uczonym, jezuitą, a jego prace miały istotny wpływ na rozwój idei, które doprowadziły do powstania sztucznej inteligencji. Początki zainteresowań jego osobą i twórczością datuje się na lata osiemdziesiąte dwudziestego wieku. Określany jest jako „ostatni człowiek, który wiedział wszystko”⁹; był osobowością prawdziwie renesansową. Poszukiwał uniwersalnego języka, rozwijał mnemotechnikę oraz kombinatorykę. Był przekonany, że słowa są rzeczywistością fundamentalną, świat bowiem został stworzony Słowem Bożym. Przedmioty są zatem upo-

⁶ Zob. A.C. Sparravigna, *Giovanni de la Fontana, Engineer and Magician*, <https://arxiv.org/abs/1304.4588>.

⁷ Por. M. Grządziel, *Filozofia, teologia i matematyka w pismach Mikołaja z Kuzy – szkic poglądów*, 2008, s. 6, https://www.chfpn.pl/new05/new05_grzadziel.pdf. Wyraźnie widoczny jest na przykład wpływ Kuzańczyka na poglądy Johanna Keplera. Zob. C. Boyer, *Historia rachunku różniczkowego i całkowitego i rozwój jego pojęć*, tłum. S. Dobrzycki, Państwowe Wydawnictwo Naukowe, Warszawa 1964.

⁸ T. Hobbes, *Lewiatan, czyli materia, forma i władza państwa kościelnego i świeckiego*, tłum. C. Znamierowski, Aletheia, Warszawa 2009, s. 118.

⁹ P. Finken, *Athanasius Kircher: The Last Man Who Knew Everything*, Routledge, New York–London 2004.

rządkowane według reguł gramatyki, która określa sposób użycia wyrazów¹⁰. Ten ontologiczny pogląd Kirchera powodował istotne implikacje: jeżeli bowiem między rzeczami a językiem istnieje ściśle przyporządkowanie, wówczas uprawniony staje się pogląd, że dokonując operacji na słowach (pojęciach), jesteśmy w stanie poznawać rzeczywistość. Uczony, stosując opracowaną przez Lullusa metodę ruchomych kół, stosował taką kombinatorykę do pojęć. Jego zamierzenia w tym zakresie uznać należy za maksymalistyczne: „Próbował wytwarzać możliwe kombinacje wszystkich skończonych alfabetów (nie tylko graficznych, lecz również matematycznych)”¹¹. Wymagało to rozwijania umiejętności szyfrowania. W tym celu Kircher projektował automaty i maszyny kodujące i dekodujące, a nawet roboty, które były w stanie poprzez specjalne tuby wydawać dźwięki¹².

Kolejnym myślicielem, który istotnie przyczynił się do powstania sztucznej inteligencji w dwudziestym wieku, był Gottfried Wilhelm Leibniz (1646-1716)¹³. Był on jednym z najwybitniejszych reprezentantów siedemnastowiecznego racjonalizmu oraz jednym z najbardziej wszechstronnych umysłów w dziejach ludzkości. Skonstruował między innymi maszynę liczącą, na której cześć zaprojektował medal z napisem: „*temu, co przewyższa człowieka*”¹⁴.

Leibniz realizował ideę języka uniwersalnego, która była przedmiotem rozważań wielu filozofów siedemnastego wieku. Jak zauważa Adam Nowaczyk: „Formalizacja języków i teorii jest produktem marzeń o zastąpieniu rozumowań rachunkiem”¹⁵. Leibniz w następujący sposób opisuje swoją koncepcję języka uniwersalnego (łac. *characteristica universalis*): „Widać stąd, że gdyby udało się wynaleźć znaki graficzne lub umowne nadające się do wyrażenia naszych myśli równie jasno i ściśle, jak arytmetyka wyraża liczby lub analiza geometryczna linie, to można byłoby w każdej dziedzinie – o ile tylko podlega rozumowaniu – zrobić wszystko, co można zrobić w arytmetyce i geometrii. Albowiem wszystkich dociekań zależnych od rozumowania dokonywano by przez przedstawienie znaków i rodzaj rachunku. [...] Nie trzeba by było tak so-

¹⁰ Por. O. B r e i d b a c h, M. T. G h i s e l i n, *Athanasius Kircher (1602-1680) on Noah's Ark: Baroque „Intelligent Design”*, „Theory: Proceedings of the California Academy of Sciences” 57(2006), nr 36, s. 999.

¹¹ T r z ę s i c k i, dz. cyt., s. 79.

¹² Por. K. V e r m e i r, *Athanasius Kircher's Magical Instruments: An Essay on „Science”, „Religion” and Applied Metaphysics*, „Studies in History and Philosophy of Science Part A” 38(2007) nr. 2, s. 363-400.

¹³ Zob. O. S t e f f e n s, *Leibniz, Lull and the Logic of Truth: Precursors of Artificial Intelligence*, Ostbayerische Technische Hochschule Regensburg, Regensburg 2020.

¹⁴ K. T r z ę s i c k i, *Leibnizjańskie inspiracje informatyki*, „Filozofia Nauki” 14(2006) nr 3(55), s. 43.

¹⁵ A. N o w a c z y k, *Wprowadzenie do filozofii matematyki*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2019, s. 63.

bie łamać głowy, jak jesteśmy do tego zmuszeni dzisiaj, a miałyby się mimo to pewność, że w naszej mocy leży dokonanie wszystkiego, co byłoby wykonalne *ex datis*”¹⁶. Choć projekt Leibniza nie doczekał się pełnej realizacji, do jego prac odwoływali się kolejni uczeni podejmujący wyzwanie skonstruowania języka uniwersalnego. Jednym z nich był Frege (1848-1925).

Wpływ Fregego na ewolucję idei sztucznej inteligencji widoczny jest przede wszystkim w jego pracach z zakresu logiki. Nie sposób oczywiście omówić tutaj, choćby w zarysie, koncepcji niemieckiego logika. Rewolucyjne znaczenie dla rozwoju logiki miała przede wszystkim jego praca z 1879 roku zatytułowana *Begriffsschrift*¹⁷. Jest ona, z jednej strony, zerwaniem z sylogistyką Arystotelesa, jako jedyną uznawaną do tej pory teorią logiczną, z drugiej – stanowi fundament pod budowę nowoczesnej logiki matematycznej. Frege odrzuca tradycję sylogistyczną, według której zdanie składa się z nazw oraz stosunków zakresowych wyznaczonych przez określone spójniki międzyzdaniami. W swoich badaniach sięga do średniowiecznej koncepcji Tomasza z Akwinu, zgodnie z którą zdanie zbudowane jest z funktorów i ich argumentów¹⁸. Program Fregego (logicyzm) był osłabioną wersją idei Leibniza. Obu filozofom towarzyszyło przekonanie o izomorficzności struktur świata, myśli oraz języka. Znaczenie badań logicznych w kontekście problematyki sztucznej inteligencji polega na tym, że logika jest pewnego rodzaju łącznikiem między językiem naturalnym a językami sformalizowanymi teorii matematycznych¹⁹.

Wkład Leibniza w powstanie sztucznej inteligencji przejawia się także w jego teorii kodowania binarnego. Idea kodu binarnego wiąże się w jego ujęciu z symboliką liczb. Według religii monoteistycznych stworzenie świata polegało na powołaniu go do istnienia przez Boga z nicości. Bóg i nicość są to dwie podstawowe zasady ontologiczne. Wychodząc od tej przesłanki, Leibniz proponuje, aby cyfra jeden stanowiła symbol odwiecznej światłości Boga, cyfra zero natomiast – nicość i ciemność. W swojej koncepcji odwołuje się także do tradycji chińskiej, którą żywo się interesował²⁰. System rachunku

¹⁶ G.W. L e i b n i z, *Przedmowa do nauki ogólnej*, w: tenże, *Wyznanie wiary filozofa*, tłum. S. Cichowicz, Wydawnictwo Naukowe PWN, Warszawa 1969, s. 71n.

¹⁷ Zob. G. F r e g e, *Ideografia. Język formalny czystego myślenia wzorowany na języku arytmetyki*, w: *Próby gramatyki filozoficznej. Antologia*, red. i tłum. K. Rotter, Wydawnictwo Uniwersytetu Wrocławskiego, Wrocław 1997.

¹⁸ Por. H. J a s t r z ę b s k i, S. C h l e b o w s k i, *Frege i Łukasiewicz o wartościach logicznych*, „Hybris” 27(2014), s. 122.

¹⁹ Por. J. M a c i a s z e k, *Od matematyki do analizy języka naturalnego. Kilka uwag o filozofii języka Gottloba Fregego*, „Językoznawstwo. Współczesne badania, problemy i analizy językoznawcze” 7(2013), s. 58.

²⁰ Por. K. T r z ę s i c k i, *Leibnizjańskie inspiracje informatyki*, „Filozofia Nauki” 14(2006) nr 3(55), s. 31.

binarnego stworzony przez Leibniza ma fundamentalne znaczenie dla informatyki. Nie był on wprawdzie pierwszym filozofem, który zainteresował się tym systemem, lecz jako pierwszy doprowadził go do postaci umożliwiającej wykorzystanie praktyczne²¹. Kazimierz Trzęsicki podkreśla znaczenie Leibniza jako właściwego twórcy systemu binarnego: „System dwójkowy został wynaleziony przez Leibniza około 1679 r., kiedy to napisał *De progressionem dyadica* [...]. W pracy tej Leibniz analizuje możliwości tego systemu oraz pokazuje, jak wykonywać podstawowe operacje arytmetyczne: dodawanie i odejmowanie, mnożenie i dzielenie. W artykule tym Leibniz pisze o maszynie działającej w oparciu o system binarny”²². Mimo że idea systemu binarnego nie znalazła uznania wśród siedemnastowiecznych uczonych, projekty i koncepcje Leibniza przetrwały i okazały się niezwykle płodne. Są one realizowane we współczesnej informatyce wykorzystującej koncepcję maszynowego „myślenia” zapisanego językiem binarnym²³.

Prace skutkujące późniejszym powstaniem sztucznej inteligencji nabrały tempa w dziewiętnastym wieku. Poza wspomnianymi już dokonaniem logicznymi Fregego kluczowe okazały się dokonania angielskiego matematyka i filozofa George’a Boole’a (1815-1864). Był on pierwszym uczonym, który – w roku 1847 – stworzył język formalny dla rozumowań logicznych. Tak zwana algebra Boole’a jest strukturą algebraiczną, która znalazła zastosowanie w informatyce oraz elektronice cyfrowej²⁴. Istota przekształceń logicznych zaproponowanych przez angielskiego matematyka nawiązuje do koncepcji Leibniza (a wcześniej Arystotelesa), że każde zdanie w sensie logicznym może być wyłącznie prawdziwe bądź fałszywe²⁵. Stosowana obecnie elektronika cyfrowa opiera się na układach logicznych, których operacje wykonywane są na prawach algebry Boole’a²⁶.

Kolejnym ważnym wydarzeniem było zbudowanie przez Alana Turinga w 1936 roku słynnej maszyny nazwanej jego imieniem. Maszyna Turinga to

²¹ Można wskazać na przykład na pitagorejczyka Filolaosa z Tarentu (ok. 480-405 p.n.e.), dla którego podstawową zasadą świata była opozycja skończone–nieskończone, czy indyjskiego poetę i matematyka Pingalę (III-II w. p.n.e.). W czasach późniejszych idee systemu binarnego rozważał Thomas Harriot (1560-1621). Por. J.W. Shirley, *Binary Numeration before Leibniz*, „American Journal of Physics” 19(1951) nr 8, s. 452-454.

²² Trzęsicki, *Leibnizjańskie inspiracje informatyki*, s. 27.

²³ Por. tamże, s. 45.

²⁴ Zob. L.M. Laita, L. de Ledesma, E. Roanes-Lozano, A. Brunori, *George Boole, a Forerunner of Symbolic Computation*, w: *Artificial Intelligence and Symbolic Computation: International Conference AISC 2000 Madrid, Spain, July 2000; Revised Papers*, red. J.A. Campbell, E. Roanes-Lozano, Springer, Berlin–Heidelberg 2001, s. 1-19.

²⁵ Por. A. Pieuch, *Zaniedbana algebra a nauczanie informatyki*, „Edukacja – Technika – Informatyka” 8(2017) nr 3, s. 289.

²⁶ Por. tamże, s. 290.

teoretyczny projekt maszyny matematycznej – automatu wykonującego samoczynnie operacje na symbolach, których celem jest znalezienie rozwiązania określonego zadania matematycznego. Założeniem, na którym opierał się pomysł Turinga, było przekonanie, że każde obliczalne (algorytmizowalne, rozstrzygalne) zadanie matematyczne może zostać rozstrzygnięte przez maszynę. Idea maszyny Turinga sformułowana została w jego artykule *On Computable Numbers with an Application to the Entscheidungsproblem*²⁷. Rewolucyjne w pomysle Turinga było wyrażone explicite przekonanie, że możliwa jest mechanizacja procesu obliczeń oraz że może tego dokonać maszyna²⁸. Maszyna Turinga była początkowo tworem wyłącznie teoretycznym. Dopiero w trakcie drugiej wojny światowej oraz po jej zakończeniu, pod wpływem doświadczeń w związku z udziałem w pracach nad łamaniem szyfrów maszyn kodujących, Turing dostrzegł w swoim metamatematycznym projekcie problemy natury praktycznej, dotyczące możliwości skonstruowania maszyny dokonującej operacji rachunkowych. Przyjmuje się, że komputery stanowią fizyczną realizację matematycznej maszyny Turinga. Sam Turing nie był oczywiście filozofem, lecz jego wkład w powstanie sztucznej inteligencji jest znaczący. Kierował się on w swoich dociekaniach filozoficzno-psychologicznymi przesłankami zaczerpniętymi z koncepcji behawioryzmu funkcjonalnego Gilberta Ryle’a (1900-1976).

Niewiele później, w roku 1943 Warren McCulloch i Walter Pitts stworzyli model sztucznych neuronów. A już w roku następnym John von Neumann i Oskar Morgenstern opracowali teorię decyzji (teorię gier)²⁹. Sformalizowali oni klasyczną teorię decyzji, co znalazło zastosowanie nie tylko w informatyce, lecz także w wielu innych naukach, między innymi w ekonomii.

Wszystkie omówione powyżej dokonania stanowiły etapy przygotowywania. Projekt sztucznej inteligencji ze sfery rozważań teoretycznych przeszedł do etapu urzeczywistnienia w połowie dwudziestego wieku. Trafnie ujął to Bruce G. Buchanan: „Jednak dopiero w ostatnim półwieczu byliśmy w stanie, jako społeczność AI, zbudować maszyny zdolne do testowania hipotez dotyczących mechanizmów myślenia i inteligentnych zachowań, a tym samym zademonstrować mechanizmy istniejące wcześniej jedynie jako teoretyczne możliwości”³⁰. Za rok narodzin sztucznej inteligencji uznaje się 1956, kiedy

²⁷ Por. A.M. Turing, *On Computable Numbers, with an Application to the Entscheidungsproblem*, „Journal of Mathematics” 58(1936) no. 5, ss. 345-363.

²⁸ Por. M. Hełmąński, *Maszyna Turinga a umysł ludzki*, „Annales Universitatis Mariae Curie-Skłodowska. Sectio I. Philosophia-Sociologia” 23(1988), s. 156.

²⁹ Zob. J. von Neumann, O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, Princeton 1944.

³⁰ B. G. Buchanan, *A (Very) Brief History of Artificial Intelligence*, „AI Magazine” 26(2005) nr 4, s. 53 (tłum. fragm. – Z.O.).

to McCarthy po raz pierwszy użył tego terminu. Odtąd obserwuje się znacznie żywsze zainteresowanie naukowców reprezentujących różne dziedziny zagadnieniem sztucznej inteligencji³¹.

SZTUCZNA INTELIGENCJA JAKO PROBLEM FILOZOFICZNY

Rozwój sztucznej inteligencji implikuje problemy natury filozoficznej. Są to głównie problemy ontologiczne, epistemologiczne oraz etyczne. Jak już wspomniano, wielu filozofów od czasów starożytnych stawiało pytania istotne z punktu widzenia metodologii sztucznej inteligencji³². Dotyczyły one między innymi natury procesów poznawczych, możliwości konstrukcji języków formalnych opisujących w ścisły sposób rzeczywistość czy idei automatyzacji rozumowań. Do wskazanych już wcześniej filozofów podejmujących tę problematykę dodać można na przykład Williama Ockhama, św. Tomasza z Akwinu czy Kartezjusza.

Wydaje się, że związek filozofii z zagadnieniami sztucznej inteligencji nie jest jedynie akcydentalny. W interesujący sposób kwestię tę ujmuje Michał Heller: „Jednym z głównych powodów nawrotu do «naturalizmu» (można wręcz mówić o «nowym naturalizmie») są ogromne postępy w naukach neurokognitywnych i w dziedzinie tak zwanej sztucznej inteligencji. Działa tu następujący mechanizm napędzający. Z jednej strony postęp w wyżej wspomnianych dziedzinach jest istotnie imponujący; z drugiej jednak strony o ludzkim umyśle wiemy ciągle bardzo mało, a to skłania do «uzupełniania» naukowych wyników filozoficznymi domysłami. Wiele neurofizjologów i specjalistów od sztucznej inteligencji, nie będąc filozofami z wykształcenia, podejmuje próby filozofowania na własną rękę. Utwierdzają ich w tym sporadyczne kontakty (bezpośrednie lub za pośrednictwem lektury) z zawodowymi filozofami, którzy bardzo często okazują się naukowo nieprzygotowani (a więc nieużyteczni) do prowadzenia takiego dialogu. Mówi się już dziś o tak zwanej neurofilozofii; powstaje ona w obszarze mało zagospodarowanym przez tradycyjną filozofię i z natury rzeczy ma ścisły związek z naukami neurokognitywnymi (niekiedy nauki te w sposób prawie niezauważalny przechodzącą w rozważania filozoficzne)”³³. Wydaje się zatem, że refleksja filozoficzna wpisana jest niejako w naturę zagadnień pojawiających się w trakcie rozwoju badań nad sztuczną inteligencją. Można nawet postawić tezę, że źródłem idei sztucznej inteligencji

³¹ Por. V. K a u l, S. E n s l i n, S. A. G r o s s, *History of Artificial Intelligence in Medicine*, „Gastrointestinal Endoscopy” 92(2020) nr 4, s. 807-812.

³² Por. M. F l a s i Ń s k i, *History of Artificial Intelligence*, w: tenże, *Introduction to Artificial Intelligence*, Springer, Cham 2016, s. 3-13.

³³ M. H e l l e r, *Chrześcijański naturalizm*, „Roczniki Filozoficzne” 51(2003) nr 3, s. 44.

jest, obok logiki matematycznej, właśnie filozofia. Pogląd ten wydaje się tym bardziej uzasadniony, jeżeli weźmie się pod uwagę, że przez wieki logika traktowana była jako część filozofii, a dokładniej narzędzie jej uprawiania.

Jednym z problemów natury ontologicznej pojawiających się przy rozważaniu zagadnienia sztucznej inteligencji jest kwestia umysłu człowieka, która, rozpatrywana w tym aspekcie, sprowadzałaby się przede wszystkim do odpowiedzi na pytanie, czy istnieje zasadnicza różnica między umysłem człowieka a inteligentną maszyną. Odpowiedź nie jest, jak mogłoby się na pierwszy rzut oka wydawać, prosta – aby odpowiedzieć na postawione pytanie, należy wcześniej ustalić, jaka jest natura umysłu, a ujmując rzecz ściślej, czy umysł jest zdeterminowany (zaprogramowany) w swoich działaniach, w jakiś sposób analogiczny do sztucznej inteligencji, która działa na podstawie stworzonego przez człowieka programu, czy też operacje umysłu cechuje jakiś rodzaj wolności. Jak dotychczas nie udało się uzyskać jednoznacznej odpowiedzi.

Jedna z przyjmowanych w naukach kognitywnych definicji umysłu stwierdza, że jest on „systemem kontrolnym określającym zachowanie się systemu przy oddziaływaniach ze złożonym, zmiennym w czasie środowiskiem”³⁴. Nie jest to jednak definicja precyzyjna, lecz jedynie relacyjne ujęcie funkcji umysłu, które niewiele, tak naprawdę, mówi o jego naturze. Urszula Żegleń zwraca uwagę na wielowymiarowy charakter koncepcji umysłu. Wielowymiarowość ta implikuje tezę o konieczności zastosowania w badaniach nad istotą umysłu podejścia interdyscyplinarnego. Żegleń utożsamia umysł z ludzkim systemem poznawczym, ale nie z mózgiem. „Przez system poznawczy rozumiem taki system, w którym zachodzą procesy poznawcze, co znaczy, że jest to system zdolny do przyjmowania informacji z zewnątrz i przekształcania ich w celu uzyskiwania (przez system) wiedzy o świecie”³⁵. Zaproponowana perspektywa przeciwstawia się zdecydowanie naturalistycznemu ujęciu umysłu. Naturalizm ten występuje w dwóch zasadniczych postaciach: (1) naturalizmu ontologicznego, głoszącego tezę o istnieniu umysłu w świecie, oraz (2) naturalizmu metodologicznego, zgodnie z którym właściwą metodą badań umysłu jest metoda naukowa. Wobec licznych trudności towarzyszących dotychczasowym próbom zdefiniowania pojęcia umysłu i naukowego opisu zjawisk mentalnych przyjąć można rozumienie umysłu jako „funkcji mózgu”³⁶. W przyjętej w niniejszym artykule perspektywie filozoficznej jego akceptacja nie rozstrzyga sporu naturalizm–antynaturalizm. Stwierdza jedynie związek umysłu z procesami neuronalnymi zachodzącymi w mózgu. Interesującą tezę wysuwa Roger Penrose,

³⁴ W. Duch, *Umysł, świadomość i działanie twórcze*, „Kognitywistyka i Media w Edukacji” 2008, nr 1-2, s. 10.

³⁵ U. Żegleń, *Wielowymiarowość umysłu*, „Filozofia Nauki” 12(2004) nr 3-4, s. 154.

³⁶ Duch, dz. cyt., s. 7.

według którego umysł (świadomość) nie daje się wyjaśnić na gruncie współczesnych neuronauk³⁷. Zdaniem brytyjskiego fizyka procesy mentalne mogą zostać wyjaśnione dopiero na poziomie kwantowym.

Wacław Branicki, referując toczącą się w powyższej sprawie dyskusję, zauważa, że założenie, że różnica między inteligentną maszyną a człowiekiem jest jedynie różnicą techniczną, związaną ze stanem rozwoju sztucznej inteligencji, zdaje się wątpliwe. „Przedstawione tutaj rozumowanie – pisze Branicki – wydaje się spójne, jednak warto pamiętać, że opiera się ono na dyskusyjnym założeniu definiującym ludzką podmiotowość jako zbiór schematów uwarunkowanych biologicznie, społecznie i językowo. Z perspektywy dezintegracji pozytywnej rysuje się inna hipoteza. Zakłada ona, że taki redukcjonistyczny obraz człowieka nie znajduje potwierdzenia w badaniach empirycznych i jest przykładem błędu logicznego typu *pars pro toto*”³⁸.

Dyskusyjne są także inne kwestie, jak zdolność zdobywania wiedzy czy wręcz jej tworzenia, charakteryzująca umysł ludzki – ale czy również sztuczną inteligencję? Pozytywnej odpowiedzi w tym względzie udziela między innymi Ryszard Tadeusiewicz, który uważa, że komputer jest w stanie generować nową wiedzę³⁹. Rodzi się pytanie o naturę tej wiedzy. Czy sztuczna inteligencja potrafi tworzyć wiedzę nową jakościowo? Czy wiedza ta ogranicza się wyłącznie do nauk formalnych, czy też maszyny zdolne są już do odkrywania wiedzy o świecie transcendentnym? Wydaje się, że obecnie maszyny tworzą jedynie wiedzę funkcjonalną. Nauki kognitywne wkroczyły dopiero na drogę swojego rozwoju. Być może ich dalszy postęp udzieli nam odpowiedzi na pytania dotyczące natury umysłu oraz pozwoli lepiej zrozumieć różnego typu procesy poznawcze.

Działanie sztucznej inteligencji rodzi także cały szereg pytań natury etycznej. Czy maszyna może stanowić moralny podmiot działania? Pojawia się zagadnienie ewentualnej odpowiedzialności moralnej sztucznej inteligencji, a także odpowiedzialności uczonych zaangażowanych w prace nad nią. Sztuczna inteligencja – jako dziedzina nauki – implikuje zarówno problemy etyczne należące do innych obszarów badań naukowych, ale też takie, które są charakterystyczne wyłącznie dla niej. Paweł Łupkowski wyróżnia trzy grupy zarysowujących się tutaj problemów: (1) problemy dotyczące odpowiedzialności uczonych pracujących na polu sztucznej inteligencji; (2) problemy odnoszące

³⁷ Zob. R. Penrose, *Nowy umysł cesarza. O komputerach, umyśle i prawach fizyki*, tłum. P. Amsterdamski, T. Lanczewski, Zysk i S-ka Wydawnictwo, Poznań 2021.

³⁸ W. Branicki, *Autentyczność osobowa w pespektywie teorii dezintegracji pozytywnej*, „*Studia Humanistyczne AGH*” 11(2012) nr 3, s. 128.

³⁹ Por. R. Tadeusiewicz, *Awangarda sztucznej inteligencji – maszyny, które potrafią same tworzyć nowe pojęcia*, w: *Pojęcia. Jak reprezentujemy i kategoryzujemy świat*, red. J. Bremer, A. Chuderski, Universitas, Kraków 2011, s. 414.

się do oceny etycznej postępu naukowo-technologicznego; (3) zagadnienie przyszłego statusu inteligentnych maszyn obdarzonych świadomością⁴⁰. Nie podejmując tego niezwykle interesującego, ale także trudnego wątku, zauważyć należy, że dalszy rozwój sztucznej inteligencji wymagał będzie wzmoczonych działań ze strony etyków. Być może niezbędne stanie się nawet utworzenie odrębnej subdyscypliny w ramach etyki.

PROBLEMATYKA ŚWIADOMOŚCI SZTUCZNEJ INTELIGENCJI

W roku 1950 Turing zaproponował sposób rozstrzygania, czy dane oprogramowanie (maszyna) jest inteligentne. Jego zdaniem za inteligentne należy uznać takie urządzenie, w odniesieniu do którego rozmawiający z nim człowiek nie jest w stanie rozpoznać, czy prowadzi dialog z maszyną, czy z innym człowiekiem. Pomijając długą historię dyskusji nad testem Turinga, zgodzić się należy z negatywną oceną uzyskanych za jego pomocą wyników. Kluczowy wydaje się argument natury filozoficznej, który w sformułowaniu Pawła Łupkowskiego brzmi następująco: „Idea tego typu argumentacji sprowadza się do wykazania, że nawet jeśli maszyna zdałaby test Turinga, to i tak nie moglibyśmy powiedzieć o niej, że jest inteligenta, ponieważ TT dostarcza z gruntu błędnej definicji posiadania inteligencji przez maszyny”⁴¹. Wspomniany autor odwołuje się do koncepcji Neda Blocka⁴², który zwraca uwagę, że test Turinga nie odróżnia kompetencji umysłu od jego zachowania. Umysł natomiast charakteryzuje jego kompetencje, a nie obserwowalne zachowania. Powyższy argument jest wersją „argumentu z drzewa konwersacji”⁴³. Zakłada on maksymalny czas trwania testu Turinga na jedną godzinę. Ma to umożliwić utworzenie wspomnianego drzewa konwersacji, zawierającego kompletną listę możliwych pytań i odpowiedzi, dających się sformułować w ciągu godziny. Inną jego odmianę stanowi argument „chińskiego pokoju” zaproponowany przez Johna R. Searle’a⁴⁴, również prowadzący do negatywnego wniosku odnośnie do możliwości przypisania inteligencji maszynie. Wydaje się zatem, że test Turinga nie jest w stanie dostarczyć przekonującej odpowiedzi na pytanie,

⁴⁰ Por. P. Ł u p k o w s k i, *Rola etyki i antropologii w rozważaniach o sztucznej inteligencji*, „Ethos” 18(2005) nr 1-2, s. 240.

⁴¹ T e n z e, *Test Turinga. Perspektywa sędziego*, Wydawnictwo Naukowe UAM, Poznań 2010, s. 32.

⁴² Por. N. B l o c k, *The Mind as the Software of the Brain*, w: *An Invitation to Cognitive Science*, t. 3, *Thinking*, red. E. Smith, D. Osherson, The MIT Press, Londyn, 1995, s. 377-425.

⁴³ Ł u p k o w s k i, *Test Turinga*, s. 32.

⁴⁴ Por. J.R. S e a r l e, *Minds, Brains, and Programs*, „Behavioral and Brain Sciences” 3(1980) nr 3, 417-457.

czy inteligentne maszyny są zdolne myśleć bądź czy cechuje je jakiś rodzaj świadomości. Sam Turing zresztą traktował sformułowany przez siebie test nie tyle jako sposób odpowiedzi na pytanie, czy maszyna jest w stanie myśleć, ile jako narzędzie do rozwiązywania skromniejszego problemu, a mianowicie, czy maszyna, biorąc udział w określonej grze, poradzi sobie równie dobrze jak człowiek⁴⁵. Pytanie o świadomość inteligentnych maszyn pozostaje zatem otwarte. Jego rozstrzygnięcie wymaga odpowiedzi na inne fundamentalne i niezwykle trudne pytanie: Czym jest świadomość?

Problematyka świadomości stanowi jedno z kluczowych zagadnień psychologii, nauk kognitywnych oraz filozofii. Świadomość to zjawisko trudne do zdefiniowania. Pytanie o jej istotę nurtowało wielu myślicieli na przestrzeni dziejów. Przedstawienie ich rozważań nie jest tutaj oczywiście możliwe. Warto jednak przynajmniej zwrócić uwagę na pewne aspekty filozoficznych analiz świadomości, które ujawniły się w trakcie trwającego wiele wieków dyskursu filozoficznego.

Znamienne jest, w jak różny sposób poszczególne szkoły filozoficzne traktowały fenomen świadomości i jak różne znaczenia przypisywały tej problematyce. Filozofowie starożytni i średniowieczni nie pisali o świadomości, lecz o duszy⁴⁶. Pojęcie świadomości odgrywało fundamentalną rolę w myśli Kartezjusza oraz wyrosłej na jej gruncie tradycji filozoficznej. Według Kartezjusza funkcją świadomości jest myślenie oraz wytwarzanie idei (łac. *cogitationes*). Myślenie, w ujęciu francuskiego filozofa, obejmuje zarówno akty poznawcze, jak i wolicjonalne oraz namiętności i emocje. Jest to najszersze możliwe ujęcie świadomości. Wymienione zjawiska psychiczne różnią się, według Kartezjusza, jedynie stopniem wysiłku jakiego wymaga ich pojawienie się w umyśle, traktowanym jako byt wytwarzający świadomość, której najwyższą formą jest samoświadomość. Świadomość zaś to jedna z dwóch podstawowych kategorii ontologicznych. Jej treścią są idee. Pojęcie „ja” zostaje przez Kartezjusza zredukowane do pojęcia świadomości, ducha, umysłu (łac. *res cogitans*)⁴⁷. Przeciwnieństwo świadomości – rzecz rozciągła, materia (łac. *res extensa*) jest maszyną, mechanizmem.

W wywodzącej się od Franza Brentana (1838-1917) tradycji podkreśla się intencjonalny charakter przeżyć świadomych, czyli fakt, że są one zawsze

⁴⁵ Por. A.M. Turing, *Computing Machinery and Intelligence*, „Mind” 59(1950) nr 236, s. 434.

⁴⁶ Por. A. Revonsuo, S. Sajama, M. Kampinen, *General Introduction: The Riddle of Consciousness*, w: *Consciousness in Philosophy and Cognitive Neuroscience*, red. A. Revonsuo, M. Kampinen, Psychology Press, New York 2014, s. 5.

⁴⁷ Por. R. Descartes, *Medytacje o pierwszej filozofii. Wraz z zarzutami uczonych mężów i odpowiedziami autora*, tłum. M. Ajdukiewicz, K. Ajdukiewicz, Państwowe Wydawnictwo Naukowe, Warszawa 1958, s. 22.

skierowane na jakiś przedmiot. Rozważania nad istotą świadomości stanowią centralny punkt filozofii ucznia Brentana – Edmunda Husserla (1859-1938). Twórca fenomenologii akceptuje w początkowej fazie swojej twórczości Brentanowską tezę o intencjonalnym charakterze przeżyć świadomych⁴⁸. W późniejszym okresie koncepcja świadomości ulega w myśli Husserla licznym modyfikacjom. W *Ideach czystej fenomenologii i fenomenologicznej filozofii* pojawia się koncepcja tak zwanej czystej transcendentalnej świadomości⁴⁹. Tak rozumiana świadomość jest bytem absolutnym stanowiącym podstawę ontologiczną świata, w którym konstytuują się sensory wszystkich przedmiotów. Prawdopodobnie pojmowana tak świadomość nie może, ex definitione, przysługiwać inteligentnym maszynom.

Rozwijana przez tradycję brentanowską koncepcja intencjonalności świadomości znalazła kontynuację we współczesnej filozofii umysłu. Zdaje się do niej nawiązywać między innymi Searle, który wskazuje, że umysłu ludzkiego nie da się sprowadzić wyłącznie do czysto poznawczych – obliczeniowych funkcji. Odrzuca on mocną wersję koncepcji sztucznej inteligencji, zgodnie z którą komputerom można przypisać świadomość. Searle uważa, że komputery nie myślą, gdyż nie posiadają nastawienia intencjonalnego: „Ich myśli nie są myślami o czymś ani też [komputery – Z.O.] nie posiadają emocji i nie przeżywają uczuć”⁵⁰. Pojęcie intencjonalności jest także kluczowe w zaproponowanym przez Searle’a słynnym argumencie „chińskiego pokoju”.

Wskazane tu wybrane filozoficzne koncepcje świadomości nie są oczywiście jedynymi, jakie można odnaleźć w europejskiej myśli filozoficznej. Świadomość jest w nich traktowana jako byt sui generis, nieredukowalny do zjawisk fizycznych. Można także odnaleźć inne perspektywy (redukcjonistyczne), na przykład ujęcie Daniela Clementa Dennetta⁵¹. Traktuje się w nich świadomość jako tak czy inaczej rozumiany wytwór procesów natury fizjologicznej zachodzących w mózgu. Istotne staje się tutaj, w kontekście rozważań nad sztuczną inteligencją, rozróżnienie tego, co „naturalne”, od tego, co „sztuczne”. Anna Latawiec zwraca uwagę, że kluczowe w tej dyskusji jest opowiedzenie się za wyróżnionymi przez Searle’a mocną bądź słabą wersją sztucznej inteligencji⁵². Uznanie pierwszej z nich prowadzi do antropomorfizowania inteligentnych

⁴⁸ Zob. E. H u s s e r l, *Badania dotyczące fenomenologii i teorii poznania*, cz. 2, tłum. J. Sidorek, Wydawnictwo Naukowe PWN, Warszawa 2000.

⁴⁹ Zob. t e n z e, *Idee czystej fenomenologii i fenomenologicznej filozofii*, t. 1, tłum. D. Gierulanka, Państwowe Wydawnictwo Naukowe, Warszawa 1975.

⁵⁰ Z. P i ą t e k, *Spory o naturę świadomości a resuscytacja*, „Humanistyka i Przyrodoznawstwo” 10(2004), s. 36.

⁵¹ Zob. D. C. D e n n e t t, *Natura umysłów. Jak zrozumieć świadomość*, tłum. W. Turopolski, Copernicus Center Press, Kraków 2021.

⁵² Por. A. L a t a w i e c, *Naturalne i sztuczne w świetle rozważań o sztucznej inteligencji*, „Studia Philosophiae Christianae” 32(1996) nr 1, s. 192.

maszyn, akceptacja drugiej może implikować negację możliwości istnienia sztucznej inteligencji w ogóle. Nie rozstrzygając, czy opisywany spór daje się sprowadzić wyłącznie do werbalnej jego postaci, uznać należy, że ustalenia terminologiczne są ważną jego częścią.

Filozoficzne ujęcia świadomości cechuje wysoki stopień ogólności oraz brak odniesień do danych empirycznych. Podstawowym problemem dotyczącym zagadnienia sztucznej inteligencji jest odpowiedź na pytanie, czy i jak świadomość może wyłaniać się z materii⁵³. I tutaj rozważania filozoficzne mogą okazać się pomocne. Negatywna odpowiedź na powyższe pytanie zdaje się wykluczać możliwość mocnej wersji sztucznej inteligencji. Próbę systematyki występujących w filozofii koncepcji świadomości przeprowadził Jakub Jonkisz. Została ona dokonana w oparciu o cztery kryteria: rodzaj epistemiczny, rząd semantyczny, typ pragmatyczny oraz stan fizjologiczny. W jej wyniku powstała tabela zawierająca dwanaście wyróżnionych typów świadomości⁵⁴. Pomimo mogących pojawiać się wątpliwości, dotyczących chociażby takiego, a nie innego wyboru kryteriów podziału, można mieć wrażenie, że próby uporządkowania problematyki świadomości podejmowane zarówno w filozofii, jak i naukach szczegółowych, należy uznać za wartościowe. Nie jest to jedyna próba systematyzacji filozoficznych koncepcji świadomości. Joseph Levine za pierwotny w różnych ujęciach świadomości uznaje podział na koncepcje materialistyczne i antymaterialistyczne⁵⁵. Pomimo dominującego w ostatnim okresie interdyscyplinarnego podejścia do problematyki świadomości, podział ten najlepiej odzwierciedla ogólne ramy prowadzonych od stuleci filozoficznych rozważań nad świadomością.

W przypadku świadomości mamy do czynienia z interesującą sytuacją poznawczą. Otóż, jak dotychczas ani filozofom, ani też przedstawicielom nauk szczegółowych podejmującym tę problematykę nie udało się zgłębić natury świadomości. Dopóty to nie nastąpi, wszelkie dyskusje dotyczące świadomości inteligentnych maszyn wydają się bezprzedmiotowe, a w każdym razie dalekie od jednoznacznych rozstrzygnięć. Uzasadnieniem powyższej tezy może być fakt istnienia wielu rozmaitych, często bardzo od siebie różnych, filozoficznych koncepcji świadomości. Nie udało się także, jak dotąd, sformułować powszechnie akceptowanej definicji świadomości.

⁵³ Por. E. H i l d t, *Artificial Intelligence: Does Consciousness Matter?*, „Frontiers in Psychology” 10(2019), s. 1535. <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01535/full>.

⁵⁴ Por. J. J o n k i s z, *Pojęcie świadomości w kognitywistyce i filozofii umysłu – próba systematyzacji*, „Filozofia Nauki” 20(2012) nr 2, s. 55.

⁵⁵ Por. J. L e v i n e, *Recent Work on Consciousness*, „American Philosophical Quarterly” 34(1997) nr 4, 379-404.

*

Sztuczna inteligencja jest jednym z najodważniejszych zamierzeń umysłu ludzkiego. Pomimo faktu, że sam termin pojawił się stosunkowo późno, bo dopiero w połowie dwudziestego wieku, wyrosła ona z wielowiekowej refleksji nad naturą poznania, a w szczególności stanowi realizację pragnienia przejawianego przez wielu myślicieli na przestrzeni wieków: ujęcia ludzkiego rozumowania w ramy logicznego formalizmu, jego algorytmizację. Idea ta traktowana początkowo w kategoriach bardziej marzenia niż naukowego projektu, w miarę doskonalenia narzędzi logicznych, wyłaniała się z mroków spekulacji, aby w drugiej połowie dwudziestego stulecia znaleźć swoją egzemplifikację w postaci inteligentnych maszyn.

Wraz z rozwojem sztucznej inteligencji rodzą się pytania natury filozoficznej. Dotyczą one przede wszystkim natury umysłu, świadomości oraz informacji. Pojawia się także pytanie o ograniczenia poznawcze uzyskiwane sztucznie za pomocą cyfrowych technik przetwarzania informacji. Mogą one być związane na przykład z odkryciem tak zwanych liczb nieobliczalnych. Są to „liczby niewymierne, których przedstawienia dziesiętnego nie może wyznaczyć, z dowolną zadaną dokładnością, żaden układ do obliczeń mechanicznych, zwany dziś maszyną Turinga”⁵⁶. Istnieje możliwość, że pewne rozumowania intuicyjne nie mogą zostać poddane mechanizacji, ponieważ nie mają one reprezentacji w liczbach obliczalnych. Dyskusja w tym temacie trwa i trudno tutaj formułować jakieś ostateczne wnioski. Warto jednak zauważyć, że wykazanie istnienia w świecie fizycznym wielkości reprezentowanych przez liczby nieobliczalne nakładałoby istotne ograniczenia natury epistemologicznej na możliwości sztucznej inteligencji. Być może pewnym wyjściem z impasu stałoby się wynalezienie komputerów kwantowych, lecz jest to wciąż kwestia przyszłości.

Jeżeli przyjmiemy, że informacja jest podstawowym elementem konstytuującym zarówno świat biologiczny, jak i technologiczny, wówczas uznać możemy ją za kategorię filozoficzną. Mieczysław Lubański zauważa, że pojęcie informacji przeszło znamiennej ewolucję. W jej wyniku ukonstytuowało się ogólnonaukowe oraz odpowiednio filozoficzne pojęcie informacji⁵⁷. Istotą wspomnianej ewolucji jest poszerzanie zakresu pojęcia informacji, z jednoczesnym zubożaniem jego treści. Staje się to bardziej zrozumiałe, gdy uświadomimy sobie, że powstanie i rozwój sztucznej inteligencji jest zjawiskiem

⁵⁶ P. Stacewicz, *Liczby nieobliczalne a granice kodowania w informatyce*, „Studia Semiotyczne” 32(2018) nr 2, s. 137.

⁵⁷ Por. M. Lubański, *Informacja i jej nośniki*, „Studia Philosophiae Christianae” 16(1980) nr 2, s. 56.

interdyscyplinarnym. Zadanie filozofów polegałoby przede wszystkim na dostarczaniu ram pojęciowych niezbędnych chociażby do zrozumienia podstawowych pojęć, takich jak: inteligencja, umysł, świadomość czy informacja.

Projekt sztucznej inteligencji jest zamysłem o charakterze technologicznym, ale wyrasta on z ducha refleksji filozoficznej nad podstawowymi zagadnieniami umysłu ludzkiego, mechanizmami jego funkcjonowania i możliwościami symulowania jego działania przez maszyny. Refleksja filozoficzna nie jest tylko źródłem procesu, który doprowadził do powstania sztucznej inteligencji. Jej rozwój niesie ze sobą nowe problemy natury filozoficznej, lecz także rzuca pewne światło na te, które rozważane są niemal od początku europejskiej filozofii. Zakres problematyki filozoficznej, jaki zarysowuje się w kontekście badań nad sztuczną inteligencją, jest bardzo szeroki. Obejmuje on między innymi filozofię umysłu, neurofilozofię, filozofię języka, etykę i epistemologię procesów poznawczych.

Jeżeli zaakceptujemy tezę, że problematyka filozoficzna wpisana jest niejako w naturę ludzką, wówczas także sztuczna inteligencja, jako wytwór człowieka, nie może być od tej problematyki wolna. Rozwój sztucznej inteligencji stawia zatem nowe wyzwania także przed filozofami. Uzasadnienie tej tezy może być dwójakie. Po pierwsze historyczne: pamiętać należy, że powstanie nauki poprzedzone było pojawieniem się refleksji filozoficznej, z której nauka wyrosła. Akceptacja poglądu, uznającego poznanie naukowe za jeden z wyznaczników ludzkiego sposobu bycia w świecie, implikuje zgodę na uznanie myśli, głoszącej, że filozofia stanowi jeden z przejawów natury ludzkiej. Drugi argument wynika z faktu, że pomimo rozwoju wiedzy naukowej i jej specjalizacji, problemy filozoficzne nie tylko nie są eliminowane z pola zainteresowań badaczy, lecz w wielu przypadkach rozwój nauki przyczynia się do wyłaniania się nowych. Istnieje także w człowieku potrzeba poddawania filozoficznym analizom szczegółowych wyników uzyskiwanych na gruncie rozmaitych dziedzin poznania naukowego. Jest tak również, jak się zdaje, w przypadku badań nad sztuczną inteligencją.

BIBLIOGRAFIA / BIBLIOGRAPHY

- Block, Ned. "The Mind as the Software of the Brain." In *An Invitation to Cognitive Science*. Vol. 3. *Thinking*. Edited by Edward E. Smith and Daniel N. Osherson. London: The MIT Press, 1995.
- Boyer, Carl. *Historia rachunku różniczkowego i całkowego i rozwój jego pojęć*. Translated by S. Dobrzycki. Warszawa: Państwowe Wydawnictwo Naukowe, 1964.
- Branicki, Wacław. "Autentyczność osobowa w pespektywie teorii dezintegracji pozytywnej." *Studia Humanistyczne AGH* 11, no. 3 (2012): 127–40.

- Breidbach, Olaf, and Michael T. Ghiselin. "Athanasius, Kircher (1602-1680) on Noah's Ark: Baroque 'Intelligent Design' Theory." *Proceedings of the California Academy of Sciences* 57, no. 36 (2006): 991–1002.
- Buchanan, Bruce G. "A (Very) Brief History of Artificial Intelligence." *AI Magazine* 26, no. 4 (2005).
- Descartes, René. *Medytacje o pierwszej filozofii: Wraz z zarzutami uczonych mężów i odpowiedziami autora*. Translated by Maria Ajdukiewicz and Kazimierz Ajdukiewicz. Warszawa: Państwowe Wydawnictwo Naukowe, 1958.
- Dennett, Daniel C. *Natura umysłów: Jak zrozumieć świadomość*. Translated by Witold Turopolski. Kraków: Copernicus Center Press, 2021.
- Duch, Włodzisław. "Umysł, świadomość i działanie twórcze." *Kognitywistyka i Media w Edukacji*, nos. 1–2 (2008): 9–37.
- Findlen, Paula. *Athanasius Kircher: The Last Man Who Knew Everything*. New York and London: Routledge, 2004.
- Flasiński, Mariusz. "History of Artificial Intelligence." In Flasiński, *Introduction to Artificial Intelligence*. Cham: Springer, 2016.
- Gottlob, Frege. "Ideografia: Język formalny czystego myślenia wzorowany na języku arytmetyki." In Franz Brentano, Gottlob Frege, and Christian Thiel, *Próby gramatyki filozoficznej: Antologia*. Edited and translated by Krzysztof Rotter. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego, 1997.
- Grządziel, Mariusz. *Filozofia, teologia i matematyka w pismach Mikołaja z Kuzy – szkic poglądów*, 2008. https://www.chfnp.pl/new05/new05_grzadzziel.pdf.
- Heller, Michał. "Chrześcijański naturalizm." *Roczniki Filozoficzne* 51, no. 3 (2003): 41–58.
- Hetmański, Marek. "Maszyna Turinga a umysł ludzki." *Annales Universitatis Mariae Curie-Skłodowska. Sectio I. Philosophia-Sociologia*, no. 23 (1988): 155–77.
- Hildt, Elisabeth. "Artificial Intelligence: Does Consciousness Matter?" *Frontiers in Psychology*, no. 10 (2019): 1535. <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01535/full>.
- Hobbes, Thomas. *Lewiatan, czyli materia, forma i władza państwa kościelnego i świeckiego*. Translated by Czesław Znamierowski. Warszawa: Fundacja Aletheia, 2009.
- Husserl, Edmund. *Badania dotyczące fenomenologii i teorii poznania*. Part 2. Translated by Janusz Sidorek. Warszawa: Wydawnictwo Naukowe PWN, 2000.
- . *Idee czystej fenomenologii i fenomenologicznej filozofii*. Vol. 1. Translated by Danuta Gierulanka. Warszawa: Państwowe Wydawnictwo Naukowe, 1975.
- Jastrzębski, Hubert, and Szymon Chlebowski. "Frege i Łukasiewicz o wartościach logicznych." *Hybris*, no. 27 (2014): 121–35.
- Jonkisz, Jakub. "Pojęcie świadomości w kognitywistyce i filozofii umysłu – próba systematyzacji." *Filozofia Nauki* 20, no. 2 (2012): 29–55.
- Kaul, Vivek, Sarah Enslin, and Seth A. Gross. "History of Artificial Intelligence in Medicine." *Gastrointestinal Endoscopy* 92, no. 4 (2020): 807–12.

- Latawiec, Anna. "Naturalne i sztuczne w świetle rozważań o sztucznej inteligencji." *Studia Philosophiae Christianae* 32, no. 1 (1996): 191–202.
- Laita, Luis M., et al. "George Boole, a Forerunner of Symbolic Computation." In *Artificial Intelligence and Symbolic Computation: International Conference AISC 2000 Madrid, Spain, July 2000 Revised Papers*. Edited by John A. Campbell and Eugenio Roanes-Lozano. Berlin and Heidelberg: Springer, 2001.
- Leibniz, Gottfried Wilhelm. "Przedmowa do nauki ogólnej." In Leibniz, *Wyznanie wiary filozofa*. Translated by Stanisław Cichowicz. Warszawa: Wydawnictwo Naukowe PWN, 1969.
- Levine, Joseph. "Recent Work on Consciousness." *American Philosophical Quarterly* 34, no. 4 (1997): 379–404.
- Lubański, Mieczysław. "Informacja i jej nośniki." *Studia Philosophiae Christianae* 16, no. 2 (1980): 55–66.
- Łupkowski, Paweł. "Rola etyki i antropologii w rozważaniach o sztucznej inteligencji." *Ethos* 18, nos. 1-2 (2005): 239–41.
- Łupkowski, Paweł. *Test Turinga. Perspektywa Sędziego*. Poznań: Wydawnictwo Naukowe UAM, 2010.
- Maciaszek, Janusz. "Od matematyki do analizy języka naturalnego: Kilka uwag o filozofii języka Gottloba Fregego." *Językoznawstwo: Współczesne badania, problemy i analizy językoznawcze* 7 (2013): 57–75.
- Marciszewski, Witold, and Roman Murawski. *Mechanization of Reasoning in a Historical Perspective*. Amsterdam and Atlanta: Editions Rodopi, 1995.
- McCarthy, John, et al. "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955." *AI Magazine* 27, no. 4 (2006): 12–14.
- Nowaczyk, Adam. *Wprowadzenie do filozofii matematyki*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego, 2019.
- von Neumann, John, and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton: Princeton University Press, 1944.
- Penrose, Roger. *Nowy umysł cesarza: O komputerach, umyśle i prawach fizyki*. Translated by Piotr Amsterdamski and Tomasz Lanczewski. Poznań: Wydawnictwo Zysk i S-ka, 2021.
- Piątek, Zdzisława. "Spory o naturę świadomości a resuscytacja." *Humanistyka i Przyrodoznawstwo* 10 (2004): 29–38.
- Piecuch, Aleksander. "Zaniedbana algebra a nauczanie informatyki." *Edukacja—Technika—Informatyka* 8, no. 3 (2017): 288–94.
- Revonsuo, Antti, Seppo Sajama, and Matti Kamppinen. „General Introduction: The Riddle of Consciousness.” In *Consciousness in Philosophy and Cognitive Neuroscience*. Edited by Antti Revonsuo and Matti Kamppinen. New York: Psychology Press, 2014.
- Searle, John R. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3, no. 3 (1980): 417–57.
- Shirley, John W. "Binary Numeration before Leibniz." *American Journal of Physics* 19, no. 8 (1951): 452–54.

- Sparavigna, Amelia Carolina. *Giovanni, de la Fontana, Engineer and Magician*. <https://arxiv.org/abs/1304.4588>.
- Stacewicz, Paweł. "Liczby nieobliczalne a granice kodowania w informatyce." *Studia Semiotyczne* 32, no. 2 (2018): 131–52.
- Stacewicz, Paweł. "Światopogląd informatyczny. Naukowe podstawy i filozoficzne perspektywy." In *Informatyka a filozofia: Od informatyki i jej zastosowań do światopoglądu informatycznego*. Edited by Paweł Stacewicz. Warszawa: Oficyna Wydawnicza Politechniki Warszawskiej, 2015.
- Steffens, Oliver. *Leibniz, Llull and the Logic of Truth: Precursors of Artificial Intelligence*, Regensburg: Ostbayerische Technische Hochschule Regensburg, 2020.
- Tadeusiewicz, Ryszard. "Awangarda sztucznej inteligencji – maszyny, które potrafią same tworzyć nowe pojęcia." In *Pojęcia: Jak reprezentujemy i kategoryzujemy świat*. Edited by Józef Bremer i Adam Chuderski. Kraków: Universitas, 2011.
- Trzęsicki, Kazimierz. "Idea sztucznej inteligencji." *Filozofia i Nauka* 8, no. 1 (2020): 69–96.
- . "Leibnizjańskie inspiracje informatyki." *Filozofia Nauki* 14, no. 3(55) (2006): 21–48.
- Turing, Alan Mathison. "Computing Machinery and Intelligence." *Mind* 59, no. 236 (1950): 433–60.
- . "On Computable Numbers, with an Application to the Entscheidungsproblem." *Journal of Mathematics* 58, no. 5 (1936): 345–63.
- Vermeir, Koen. "Athanasius Kircher's Magical Instruments: An Essay on 'Science', 'Religion' and Applied Metaphysics." *Studies in History and Philosophy of Science Part A* 38(2007) no. 2: 363–400.
- Żegleń, Urszula. "Wielowymiarowość umysłu." *Filozofia Nauki* 12, no. 3-4 (2004): 153–70.

ABSTRAKT /ABSTRACT

Zbigniew ORBIK – W poszukiwaniu filozoficznych źródeł problematyki sztucznej inteligencji

DOI 10.12887/36-2023-3-143-05

Celem artykułu jest poszukiwanie filozoficznych korzeni idei sztucznej inteligencji. Dokonano w nim przeglądu długiego procesu rozwoju pojawiających się w myśli filozoficznej już od czasów starożytnych idei, które uznać można za antecedeny współczesnej koncepcji sztucznej inteligencji. Wprawdzie sztuczna inteligencja rozumiana jest obecnie jako rodzaj inteligencji, którą posiadają maszyny będące w stanie symulować procesy myślenia charakterystyczne dla człowieka i jest dziedziną zaliczaną do informatyki i nauk kognitywnych, okazuje się jednak, że ma ona swoje źródło w refleksji filozoficznej. Koncepcja ta rodzi także wiele problemów natury filozoficznej, które w niniejszym tekście zostały, z oczywistych względów, jedynie zasygnalizowane. Są to zagadnienia

obejmujące swoim zakresem niemal cały przekrój problematyki tradycyjnie zaliczanej do filozofii. Dotyczą one między innymi struktury rozumowań, możliwości ich automatyzacji, natury procesów poznawczych, możliwości konstrukcji języków formalnych czy ontologicznej natury umysłu. Jednym z najważniejszych jest zagadnienie świadomości. W artykule zastosowano analizę wybranej, dostępnej literatury dotyczącej historii sztucznej inteligencji oraz jej filozoficznych aspektów. Oprócz przeglądu literatury autor stosuje metodę analizy i konstrukcji logicznej. Rozwój sztucznej inteligencji implikuje także konieczność dalszych badań filozoficznych nad naturą umysłu, istotą świadomości, procesów poznawczych oraz licznych kwestii etycznych. Filozoficzna refleksja nad sztuczną inteligencją wpisuje się w starożytny postulat samopoznania człowieka.

Słowa kluczowe: nauki kognitywne, sztuczna inteligencja, umysł, świadomość

Kontakt: Katedra Stosowanych Nauk Społecznych, Wydział Organizacji i Zarządzania, Politechnika Śląska, ul. Roosevelta 26-28, 41-800 Zabrze
E-mail: zbigniew.orbik@polsl.pl
Tel. 721 433 549
ORCID: 0000-0001-7092-1567

Zbigniew ORBIK, In Search of Philosophical Sources of the Problem of Artificial Intelligence

DOI 10.12887/36-2023-3-143-05

The aim of the article is to inquire into the philosophical origins of the idea of artificial intelligence. The author provides a review of the long process of developing ideas which have emerged in philosophical thought since antiquity and can be regarded as antecedents to the modern concept of artificial intelligence. Currently understood as a type of intelligence possessed by machines capable of stimulating human-specific thinking processes and, as a research area, considered as part of computer science and cognitive sciences, artificial intelligence has its source in philosophical reflection. Also, the concept of artificial intelligence poses numerous philosophical problems, which, for obvious reasons, were only mentioned in the article. They are issues that cover almost the entire cross-section of problems traditionally considered as philosophical and concern, among others, the structures of reasoning, the possibilities of their automation, the nature of cognitive processes, the possibilities of constructing formal languages or the ontological nature of mind. One of the most important question is that of consciousness. The article analyzes selected available literature on the history of artificial intelligence and its philosophical aspects. Apart from the literature review, the author uses the method of analysis and logical construction. The development of artificial intelligence also implies the need for further philosophical research on the nature of mind, the essence of consciousness, cognitive processes, and a vast range of emerging ethical issues.

Philosophical reflection on artificial intelligence meets the ancient postulate of human self-knowledge.

Keywords: cognitive science, artificial intelligence, mind, consciousness

Contact: Department of Applied Social Sciences, Faculty of Organization and Management, Silesian University of Technology, ul. Roosevelta 26-28, 41-800 Zabrze, Poland

E-mail: zbigniew.orbik@polsl.pl

Phone: +48 721 433 549

ORCID: 0000-0001-7092-1567

Ks. Marcin FERDYNUS

CZY MAMY PRAWO OSWOIĆ SZTUCZNĄ INTELIGENCJĘ?

Zasadniczym celem prowadzonych analiz jest wskazanie racji przemawiających za tym, że człowiek ma moralne prawo, by oswoić sztuczną inteligencję. Słowo „oswoić” w tym kontekście będzie rozumiane nie tyle jako przyzwyczaić się czy też zaprzyjaźnić się z nią, ile podporządkować ją człowiekowi. Racje na rzecz głoszonej tu tezy będą poszukiwane w dwóch głównych obszarach. Pierwszy skoncentruje się wokół negatywnych konsekwencji, które niesie ze sobą sztuczna inteligencja. Drugi zaś będzie wiązał się ze znalezieniem odpowiedzi na pytanie, czy sztuczną inteligencję można nazwać podmiotem moralnym.

Sztuczna inteligencja (SI) jest obecna w wielu wymiarach codziennego życia, choćby wtedy, kiedy przeglądamy bazy danych w poszukiwaniu różnych informacji w wyszukiwarce Google. SI jest coraz doskonalsza, pokonała mistrzów świata w gry Jeopardy!, Go oraz mistrzów klasycznych gier w szachy, backgammona¹, a ostatnio – jak donoszą portale społecznościowe – w brydża. W książce pod tytułem *Świadome maszyny. Sztuczna inteligencja i projektowanie umysłów* Susan Schneider sugeruje, że nie istnieje jeszcze SI o ogólnym zastosowaniu, a więc taka, która potrafi prowadzić inteligentną rozmowę, posiada zdolność integrowania idei dotyczących różnych zagadnień czy może myśleć lepiej niż człowiek. Tego typu SI pojawia się w filmach, takich jak na przykład *Ex Machina* lub *Ona*, może zatem sprawiać wrażenie czegoś rodem z science fiction². Schneider jednak przewiduje, że perspektywa pojawienia się SI, która będzie potrafiła łączyć informacje z różnych dziedzin oraz wykazywać się zdrowym rozsądkiem, nie jest wcale odległą³. Vincent C. Müller i Nick Bostrom podkreślają z kolei, że rozwój SI sprawi, że pewne zawody wykonywane obecnie przez ludzi staną się przestarzałe i niepotrzebne, inne zaś zostaną przez nią zastąpione. Sugerują oni, że SI będzie wykonywała niektóre zawody i funkcje co najmniej tak samo dobrze lub nawet lepiej niż ludzie⁴. Ponadto Bostrom przewiduje, że SI przekształci się w superinteligencję,

¹ Por. K. K o w a l c z e w s k a, *Sztuczna inteligencja na wojnie. Perspektywa międzynarodowego prawa humanitarnego konfliktów zbrojnych*, Wydawnictwo Naukowe Scholar, Warszawa 2021, s. 30n.

² Por. S. S c h n e i d e r, *Świadome maszyny. Sztuczna inteligencja i projektowanie umysłów*, tłum. J. Bednarek, Wydawnictwo Naukowe PWN, Warszawa 2021, s. 19.

³ Por. tamże, s. 20.

⁴ Por. V.C. M ü l l e r, N. B o s t r o m, *Future Progress in Artificial Intelligence: A Survey of Expert Opinion*, w: *Fundamental Issues of Artificial Intelligence*, red. V.C. Müller, Synthese Library, Springer, Berlin 2016, s. 553-571.

czyli w syntetyczną inteligencję, która będzie przewyższała najinteligentniejszych ludzi pod każdym względem, także rozumowania zdroworozsądkowego i umiejętności społecznych. Jego zdaniem niekontrolowany rozwój SI okaże się katastrofalny w skutkach, doprowadzi bowiem do zniszczenia ludzkości⁵.

Przedstawione tu w zarysie przewidywania dotyczące natury SI są omawiane przez przedstawicieli różnych specjalizacji. Niniejszy tekst stanowi próbę włączenia się do tej dyskusji w skromnym zakresie, obejmującym poszukiwanie odpowiedzi na pytanie tytułowe. Zasadniczym celem prowadzonych analiz jest wskazanie racji przemawiających za tym, że człowiek ma moralne prawo, by oswoić SI. Słowo „oswoić” w tym kontekście będzie rozumiane nie tyle jako przyzwyczaić się czy też zaprzyjaźnić się z SI, ile podporządkować ją człowiekowi⁶. Racje na rzecz głoszonej tu tezy będą poszukiwane w dwóch głównych obszarach. Pierwszy skoncentruje się wokół negatywnych konsekwencji, które niesie ze sobą SI. Drugi zaś będzie wiązał się z poszukiwaniem odpowiedzi na pytanie, czy SI można nazwać podmiotem moralnym. Od rozstrzygnięcia zwłaszcza tej ostatniej kwestii zależeć będzie, czy uprawomocnione jest roszczenie do podporządkowania SI człowiekowi. Analizy rozpoczną się od przedstawienia wybranych koncepcji SI, które będą stanowiły tło dla prowadzonej w artykule dyskusji.

CZYM JEST (BĘDZIE?) SZTUCZNA INTELIGENCJA?

W literaturze przedmiotu pojawiają się różne odpowiedzi na pytanie, czym jest SI. Selmer Bringsjord i Naveen Sundar Govindarajulu, autorzy hasła „Artificial Intelligence” („Sztuczna inteligencja”), opublikowanego w *Stanford Encyclopedia of Philosophy*, sugerują, że SI to dziedzina poświęcona tworzeniu sztucznych zwierząt (lub przynajmniej sztucznych stworzeń, które w odpowiednich kontekstach wydają się zwierzętami) oraz sztucznych osób (lub przynajmniej sztucznych stworzeń, które w odpowiednich kontekstach wydają się osobami)⁷. Zdaniem autorów tak sformułowane cele sprawiają, że SI stanowi przedmiot zainteresowania zwłaszcza filozofów, którzy starają się wykazać, że cele te są lub nie są w rzeczywistości osiągalne. Ponadto Bringsjord i Govindarajulu podkreślają, że

⁵ Zob. N. B o s t r o m, *Superinteligencja. Scenariusze, strategie, zagrożenia*, tłum. D. Konowrocka-Sawa, Wydawnictwo Helion, Gliwice 2016.

⁶ Przez frazę „podporządkować SI człowiekowi” można rozumieć postulat wzmocnienia regulacji ograniczających rozwój i stosowanie SI na podstawie następujących przesłanek: (1) ze względu na dynamikę postępu technologicznego w dziedzinie SI dalsza społecznie niekontrolowana ekspansja SI może stanowić poważne zagrożenie zarówno dla ludzkości jako całości, jak i dla poszczególnych jednostek, (2) nie istnieją wystarczająco silne racje natury moralnej, które sprzeciwiają się instrumentalizacji SI przez człowieka. Ten komentarz zawdzięczam jednemu z recenzentów.

⁷ Zob. S. B r i n g s j o r d, N. S. G o v i n d a r a j u l u, hasło „Artificial Intelligence”, w: *The Stanford Encyclopedia of Philosophy*, red. E.N. Zalta, <https://plato.stanford.edu/entries/artificial-intelligence/#WhatExacAI>.

wiele podstawowych formalizmów i technik stosowanych w SI wprost wywodzi się z filozofii: logika pierwszego rzędu i jej rozszerzenia, logiki intensjonalne, nadające się do modelowania postaw doksastycznych i rozumowania deontycznego, logika indukcyjna, teoria prawdopodobieństwa i rozumowanie probabilistyczne, praktyczne rozumowanie oraz planowanie. Z tego między innymi powodu niektórzy naukowcy prowadzą badania dotyczące SI w ramach filozofii⁸.

Wśród odpowiedzi na pytanie, czym jest SI, nie brakuje takich, które starają się określić SI w kontekście socjologicznym oraz informatycznym. Kinga Bączyk-Rozwadowska sugeruje, że z perspektywy socjologicznej SI to „zdolność maszyny do naśladowania ludzkiej inteligencji – system, który pozwala na wykonywanie zadań wymagających procesu uczenia się i uwzględniania nowych okoliczności w toku rozwiązywania danego rodzaju problemów”⁹. System ten może działać w sposób autonomiczny, jak również wchodzić w interakcje z otoczeniem. Z kolei perspektywa informatyczna zakłada, że SI jest systemem komputerowym, który najpierw analizuje duże ilości danych (na przykład w celu kategoryzacji i znalezienia w tych danych powtarzalności), a następnie, w oparciu o te dane, rozwiązuje zadania i podejmuje decyzje. System ten potrafi „nie tylko się «uczyć» na podstawie gromadzonych i analizowanych danych, ale także kontynuować tę naukę podczas swojego działania, w ramach którego sposób podejmowania decyzji jest stale optymalizowany, a baza danych i wiedzy – systematycznie rozszerzana”¹⁰. Perspektywę informatyczną SI zdaje się potwierdzać definicja zaproponowana przez Ryszarda Tadeusiewicza. Autor ten uważa, że „ze sztuczną inteligencją mamy do czynienia wtedy, gdy maszyna (komputer albo elektronicznie sterowane urządzenie: robot, autonomiczny pojazd, samoorganizująca się sieć połączeń) przejawia zachowania, które obserwowane u człowieka powodowałyby, że bylibyśmy skłonni je uznać za skutek jego inteligencji”¹¹.

Wydaje się, że jedną z najbardziej interesujących odpowiedzi na pytanie, czym jest SI, udzielają Stuart Russell i Peter Norvig. Wskazują bowiem na osiem definicji SI, uporządkowanych według czterech kategorii i dwóch wymiarów (zob. tabela 1). Definicje umieszczone na górze tabeli dotyczą pro-

⁸ Zob. tamże.

⁹ K. Bączyk-Rozwadowska, *Odpowiedzialność cywilna za szkody wyrządzone w związku z zastosowaniem sztucznej inteligencji w medycynie*, „Przegląd Prawa Medycznego” 8(2021) nr 3-4, s. 6.

¹⁰ Tamże.

¹¹ Por. R. Tadeusiewicz, *Archipelag sztucznej inteligencji. Część I*, „Napędy i Sterowanie” 22(2020) nr 12, s. 27. Autor ten podkreśla, że sensowne mówienie o SI wymaga uwzględnienia dwóch sposobów (metod) jej rozwijania: (a) metod całościowych (głównie są to metody heurystyczne, na przykład: sieci neuronowe, algorytmy genetyczne, logika rozmyta, kopiowanie natury) oraz (b) metod symbolicznych (między innymi automatyczne dowodzenie twierdzeń, gry, systemy ekspertowe, język PROLOG). Szczegółowa charakterystyka wymienionych metod por. tamże, s. 28-40.

cesów myślowych i rozumowania, natomiast te umiejscowione na dole tabeli odnoszą się do zachowania. Definicje znajdujące się po lewej stronie tabeli wskazują, że celem SI jest dopasowanie do ludzkiej wydajności, z kolei te po prawej stronie tabeli wskazują, że celem SI jest dopasowanie do idealnej miary wydajności, zwanej racjonalnością. Ich zdaniem system jest racjonalny, jeśli czyni „właściwą rzecz”, biorąc pod uwagę to, co wie¹².

Tabela 1.

Myślenie ludzkie „Nowy, ekscytujący wysiłek, aby komputery myślały [...] m a s z y n y z u m y s ł a m i, w pełnym i dosłownym tego słowa znaczeniu”¹². „[Automatyzacja¹³] czynności, które kojarzymy z ludzkim myśleniem, czynności takie jak: podejmowanie decyzji, rozwiązywanie problemów, uczenie się”¹⁴.	Myślenie racjonalne „Badanie zdolności umysłowych za pomocą modeli obliczeniowych”¹⁵. „Badania nad metodami obliczeniowymi, które umożliwiają postrzeganie, rozumowanie i działanie”¹⁶.
Działanie ludzkie „Sztuka tworzenia maszyn, które wykonują funkcje, które wymagają inteligencji, kiedy są wykonywane przez ludzi”¹⁷. „Dziedzina nauki o tym, jak sprawić, by komputery czyniły rzeczy, które aktualnie lepiej wykonują ludzie”¹⁸.	Działanie racjonalne „Inteligencja obliczeniowa jest to nauka dotycząca projektowania inteligentnych agentów”¹⁹. „SI [...] dotyczy inteligentnego zachowania artefaktów”²⁰.

Źródło: S. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach*, Prentice Hall, Saddle River 2009, s. 2.

¹² Por. S. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach*, Prentice Hall, Saddle River 2009, s. 1n. (Jeśli nie podano inaczej, tłumaczenie fragmentów prac obcojęzycznych – M.F.) Russell i Norvig podkreślają też, że podejście skoncentrowane na człowieku musi być częściowo nauką empiryczną, obejmującą obserwacje i hipotezy, które dotyczą ludzkiego zachowania, natomiast podejście racjonalistyczne musi obejmować połączenie matematyki i inżynierii. Por. tamże.

¹³ J. Haugeland, *Artificial Intelligence: The Very Idea*, MIT Press, Cambridge 1985, s. 2.

¹⁴ Por. Russell, Norvig, dz. cyt., s. 2.

¹⁵ R. Bellman, *An Introduction to Artificial Intelligence: Can Computers Think?*, Boyd & Fraser Publishing Company, San Francisco 1978; cyt. za: Russell, Norvig, dz. cyt., s. 5.

¹⁶ E. Charniak, D. McDermott, *Introduction to Artificial Intelligence*, Addison-Wesley, Reading 1985.

¹⁷ P.H. Winston, *Artificial Intelligence*, Addison-Wesley, Reading 1992.

¹⁸ R. Kurzweil, *The Age of Intelligent Machines*, MIT Press, Cambridge 1990.

¹⁹ E. Rich, K. Knight, *Artificial Intelligence*, McGraw-Hill, New York 1991.

²⁰ D. Poole, A. Mackworth, R. Goebel, *Computational Intelligence: A Logical Approach*, Oxford University Press, New York 1998.

²¹ N.J. Nilsson, *Artificial Intelligence: A New Synthesis*, Morgan Kaufmann, San Francisco 1998.

Zaproponowana przez Russella i Norviga klasyfikacja zakłada, że SI powinna być zdefiniowana pod kątem jej celów. Definicja SI przybiera zatem następującą ogólną postać: „SI jest dziedziną, która ma na celu budowanie”²². Biorąc pod uwagę wymienione w tabeli cztery kategorie oraz ogólną postać definicji SI: „SI jest dziedziną, która ma na celu budowanie”..., można wyróżnić cztery główne definicje SI. Po pierwsze, SI jest dziedziną, która ma na celu budowanie systemów, które myślą jak ludzie. Po drugie, SI jest dziedziną, która ma na celu budowanie systemów, które działają jak ludzie. Po trzecie, SI jest dziedziną, która ma na celu budowanie systemów, które myślą racjonalnie. Po czwarte, SI jest dziedziną, która ma na celu budowanie systemów, które działają racjonalnie. Przedstawione tu definicje podsumowuje tabela 2.

Tabela 2.

	Oparte na człowieku	Oparte na racjonalności
Oparte na rozumowaniu	Systemy, które myślą jak ludzie.	Systemy, które myślą racjonalnie.
Oparte na zachowaniu	Systemy, które działają jak ludzie.	Systemy, które działają racjonalnie.

Źródło: S. Bringsjord, N.S. Govindarajulu, hasło „Artificial Intelligence”, w: *The Stanford Encyclopedia of Philosophy*, red. E.N. Zalta, <https://plato.stanford.edu/entries/artificial-intelligence/#WhatExacAI>.

Russell i Norvig skłaniają się ku opinii, że SI to badanie racjonalnych agentów odbierających percepcje ze środowiska i wykonujących działania. Agent jest czymś, co działa, przy czym działanie charakteryzuje się autonomicznością, postrzeganiem swojego środowiska, dostosowywaniem się do zmian, jak również tworzeniem i realizowaniem celów. Racjonalny agent działa w taki sposób, by osiągnąć najlepszy rezultat lub najlepszy oczekiwany rezultat. Innymi słowy, idealny, racjonalny agent podejmuje najlepsze możliwe działanie w danej sytuacji²³. Russell i Norvig postrzegają SI w kontekście bytów (systemów) racjonalnie działających²⁴.

Przytoczone definicje SI są najczęściej formułowane pod kątem celów, posiadanych umiejętności oraz poziomu autonomii²⁵. Na tej podstawie doko-

²² Zob. Bringsjord, Govindarajulu, dz. cyt.

²³ Por. Russell, Norvig, dz. cyt., s. 4.

²⁴ Por. tamże, s. 30.

²⁵ W literaturze wskazuje się na różne poziomy autonomiczności SI: (1) program komputerowy, który wykonuje wcześniej zaprogramowane polecenia (brak autonomiczności, w zasadzie trudno jest mówić o SI); (2) przetwarzanie danych i prezentowanie rezultatów z wykorzystaniem uczenia się (system nie podejmuje żadnych decyzji); (3) przetwarzanie danych i prezentowanie rezultatów z wykorzystaniem zdolności uczenia się (system proponuje działania do podjęcia, które mogą być

nuje się dystynkcji między słabą SI i silną SI (ogólna SI). Słaba SI to zdolność maszyny (systemu) do przetwarzania informacji, a także zdolność do funkcjonowania w sposób podobny do inteligencji człowieka, niemniej jednak z określonym stopniem kontroli ze strony konstruktora (model koneksjonistyczny). Natomiast silna SI to zdolność systemu do myślenia w sposób niesymulowany (model klasyczny). Innymi słowy, silna SI ma naśladować umysł ludzki i być wyposażona w świadomość swojego istnienia²⁶. Wszystkie obecnie znane systemy opierają się na słabej SI, natomiast nie ma pewności co do tego, czy kiedykolwiek powstanie silna SI²⁷. Warto jednak podkreślić, że wśród autorów badających naturę SI nie brakuje opinii wskazujących zarówno na nieuniknioną powstania silnej SI, jak i koncepcji mówiących o tym, że silna SI przekształci się w ogólną nadludzką inteligencję (superinteligencję)²⁸. Superinteligencja to rodzaj inteligencji, która będzie miała przewagę nad ludźmi w każdej dziedzinie. W książce pod tytułem *Superinteligencja. Scenariusze, strategie, zagrożenia* Bostrom definiuje superinteligencję jako „każdy umysł, który pod względem zdolności poznawczych znacznie przewyższa człowieka w dosłownie w każdej dziedzinie zainteresowań”²⁹. Tę ogólną definicję Bostrom dookreśla, wskazując na trzy formy superinteligencji: (1) superinte-

zaakceptowane lub odrzucone); (4) przetwarzanie danych i prezentowanie rezultatów z wykorzystaniem zdolności uczenia się oraz wykonywanie działań. Na tym ostatnim etapie wskazuje się na dwie możliwości dotyczące procesu decyzyjnego: (a) człowiek musi zatwierdzić działanie lub (b) SI zarządza całością procesu decyzyjnego w sposób autonomiczny. Wiele wskazuje na to, że obecnie osiągnięto poziom (a), natomiast nie osiągnięto jeszcze poziomu (b) (por. O. Siłarski, M. Siłarski, *Cyfrowe wspomaganie decyzji medycznych w świetle prawa karnego*, „Przegląd Prawa Medycznego” 8(2021) nr 3-4, s. 4ln.).

²⁶ Por. Bączyski - Rozwadowska, dz. cyt., s. 7. Por. też: M. Bieroński, *Etyczne i moralne wyzwania związane ze stosowaniem sztucznej inteligencji*, „Kieleckie Studia Teologiczne” 2020, nr 19, s. 9.

²⁷ Por. Schneider, dz. cyt., s. 19; Bączyski - Rozwadowska, dz. cyt., s. 7.

²⁸ Transhumanista Ray Kurzweil prognozuje, że silną SI uda się skonstruować do 2030, a osobliwość do roku 2045. Uważa też, że niebiologiczna inteligencja stworzona w 2045 roku będzie miliardy razy potężniejsza niż cała dzisiejsza ludzka inteligencja (por. R. Kurzweil, *Nadchodzi osobliwość. Kiedy człowiek przekroczy granice biologii*, tłum. E. Chodkowska, A. Nowosielska, Kurhaus Publishing, Warszawa 2013, s. 134). Z kolei Vernor Vinge pisze, że będzie zaskoczony, jeśli stworzenie inteligencji technologicznej przekraczającej możliwości ludzkiej inteligencji nie dokona się do roku 2030 (por. V. Vinge, *Technological Singularity*, w: *The Transhumanist Reader. Classical and Contemporary Essays on the Science, Technology, and Philosophy of the Human Future*, red. M. More, N. Vita-More, Wiley-Blackwell, Oxford 2013, s. 366).

²⁹ Bostrom, *Superinteligencja*, s. 45. Warto dodać, że pod względem poznawczym systemy bazujące na SI już dzisiaj dystansują najinteligentniejszych ludzi, a pod względem decyzyjno-operacyjnym systemy autonomiczne, posiadające zdolność uczenia się (zdolne do samodzielnej zmiany reguł i algorytmów, na których bazują ich procesy decyzyjne), tak doskonale imitują człowieka, że z perspektywy zewnętrznego obserwatora (perspektywa fenomenologiczna), granica między człowiekiem a maszyną staje się coraz mniej uchwytna. Tę uwagę zawdzięczam jednemu z recenzentów.

ligencja szybka – system, który potrafi dokonać wszystkiego tego, co ludzki umysł, lecz znacznie szybciej (na przykład superinteligencja będzie potrafiła przeczytać książkę w kilka sekund i napisać pracę doktorską w ciągu popołudnia), (2) superinteligencja zbiorowa – system złożony z dużej liczby mniejszych form rozumnych, który znacznie przewyższa jakikolwiek współczesny system poznawczy w wielu bardzo ogólnych dziedzinach (poszczególne formy rozumne nie muszą być superinteligentne, ale rezultat ich wspólnego działania znacznie przewyższa inteligencję pojedynczego człowieka), (3) superinteligencja jakościowa – system, który jest przynajmniej równie szybki jak umysł ludzki, a przy tym znacznie inteligentniejszy (superinteligencja jakościowa jest to inteligencja co najmniej w takim stopniu przewyższająca inteligencję ludzką, w jakim inteligencja ludzka góruje nad inteligencją słoni, delfinów lub szympanów)³⁰. Bostrom uważa, że wszystkie trzy formy superinteligencji mogą ze sobą współistnieć, a także określać cele, które będą realizować. Stawia on następującą tezę (tak zwana teza o ortogonalności): „inteligencja i motywacja są ortogonalne w tym sensie, że z zasady prawie każdy poziom inteligencji można połączyć z prawie każdym celem działania”³¹. Innymi słowy, to, że superinteligencja jest inteligentna nie oznacza, że będzie stawiać sobie mądre cele. Wszystkie możliwości superinteligencji mogą zostać zaangażowane w absurdalny projekt³². W szeroko cytowanym eksperymencie myślowym Bostrom odwołuje się do przykładu superinteligencji prowadzącej fabrykę spinaczy. Jej celem ostatecznym jest ich wytwarzanie. Chcąc wykonać spinacze do papieru, superinteligencja mogłaby wykorzystać do jego realizacji całą ziemską materię, likwidując przy okazji całe życie biologiczne, w tym atomy tworzące ludzkie ciała³³. Bostrom sugeruje, że nie możemy zakładać, iż superinteligencja będzie aprobować wartości łączone stereotypowo z wiedzą i rozwojem intelektualnym ludzkości (na przykład: naukowa ciekawość, altruizm, bezinteresowność, kontemplacja, upodobanie do wyrafinowanej kultury, życiowych przyjemności i tak dalej). Zdaniem badacza nie możemy też zakładać, że superinteligencja, która stawia sobie za cel produkcję spinaczy do papieru, ograniczy swoje działania w taki sposób, żeby nie naruszyć interesów ludzkości. Obierając taki cel ostateczny, superinteligencja będzie mogła eliminować potencjalne zagrożenie dla samej siebie i dla swojej hierarchii

³⁰ Por. tamże, s. 88-94.

³¹ Tamże, s. 162. Por. też: B o s t r o m, *The Superintelligent Will: Motivation and Instrumental Rationality in Advanced Artificial Agents*, „Minds and Machines” 22(2012) nr 2, s. 71-85; S. A r m s t r o n g, *General Purpose Intelligence: Arguing the Orthogonality Thesis*, „Analysis and Metaphysics” 2013, nr 12, s. 68-84.

³² Por. S c h n e i d e r, dz. cyt., s. 163.

³³ Por. B o s t r o m, *Superinteligencja*, s. 184-186.

wartości; będzie mogła eliminować istoty ludzkie lub wykorzystać je jako zasób fizyczny³⁴.

Mając na uwadze to, co zostało do tej pory powiedziane, można z pewnym uproszczeniem przyjąć, że wysuwane przez różnych autorów opinie o SI wskazują zasadniczo na trzy główne jej typy: (1) słaba SI (obecnie stosowana), (2) silna SI (taka, która będzie w stanie dorównać człowiekowi), (3) superinteligencja (taka, która przewyższy możliwości człowieka). W dalszej kolejności, spoglądając na to, czym jest SI i czym może się stać, trzeba się zastanowić nad pytaniem o realne i prawdopodobne, zarówno indywidualne, jak i społeczne, negatywne konsekwencje, które wiążą się z (samo)rozwojem SI. Wskazanie przynajmniej na niektóre z nich może stanowić podstawę do tego, by twierdzić, że SI powinna być podporządkowana człowiekowi.

OBAWY I NEGATYWNE KONSEKWENCJE

Obawy o krótkofalowe i długofalowe negatywne konsekwencje dotyczące zastosowania SI nie są bezpodstawne. Niektóre mają charakter bardzo ogólny, inne zaś konkretny. Wydaje się, że jednym z głównych zagrożeń związanych z SI jest bezrobocie technologiczne³⁵. Wiele wykonywanych ręcznie i niewymagających specjalnych kwalifikacji prac biurowych oraz tych w fabrykach zniknie, a wraz z nimi ulegną likwidacji stanowiska pracy. Podobny los spotka inne zawody. Zostaną one zautomatyzowane³⁶. Do wzrostu bezrobocia przyczynią się też pojazdy autonomiczne. Ponadto niektóre prace, na przykład w zakresie prawa i księgowości, obejmujące czasochłonne badania nad precedensami i szczegółowymi przepisami, również zostaną zastąpione przez systemy SI. Podczas gdy jedne zawody prawnicze okażą się zbędne, inne zyskają na SI, ponieważ powstaną nowe kazusy prawne. Dobrym przykładem może być pytanie, kto powinien odpowiadać za negatywne konsekwencje wynikające z zastosowania SI: użytkownik, sprzedawca, programista, firma produkująca SI? Niewykluczone, że w niedalekiej przyszłości pojawią się takie sytuacje, w których będzie można pozwać człowieka za nieużywanie SI³⁷. Spory sądowe mogą okazać się nieuniknione, jeśli tylko zostanie udowodnione, że system SI był wysoce wiarygodny, a człowiek nie skorzystał z niego, doprowadzając do negatywnych konsekwencji (na przykład lekarz diagnozujący i leczący

³⁴ Por. tamże, s. 174.

³⁵ Por. M. B o d e n, *Sztuczna inteligencja. Jej natura i przyszłość*, tłum. T. Sieczkowski, P. Fulmański, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2020, s. 173.

³⁶ Zob. R. B a l d w i n, *The Globotics Upheaval: Globalisation, Robotics, and the Future of Work*, Oxford University Press, New York 2019.

³⁷ Por. B o d e n, dz. cyt., s. 174.

pacjenta nie skorzystał z wiarygodnej SI i doprowadził do pogorszenia stanu jego zdrowia).

Rozwój SI stanowi także zagrożenie związane z redukcją miejsc pracy w sektorze usług. Przykładowo, edukacja coraz bardziej otwiera się na internetowe pomoce naukowe związane z SI. Masowe Otwarte Kursy Online (ang. Massive Open Online Courses) oferują wykłady znanych nauczycieli akademickich³⁸, pogarszając tym samym warunki pracy wielu „tradycyjnych” nauczycieli. Z kolei zmiany demograficzne na świecie zachęcają firmy technologiczne do prowadzenia badań w obszarze „robotów-opiekunów” dla osób starszych, czy też „robotów-niań” dla dzieci³⁹. Wiele wskazuje na to, że w przypadku redukcji zatrudnienia oraz przejścia w tryb powszechnego używania systemów SI, będą musiały powstać nowe miejsca pracy. Wątpliwe jednak jest to, czy będą one równoważne pod względem liczby, dostępności edukacyjnej i czy da się z nich utrzymać⁴⁰.

Nie wchodząc w szczegóły dotyczące zawłaszczania przez SI różnych obszarów ludzkiej aktywności (pracy), warto zauważyć, że dotykamy w tym miejscu fundamentalnej kwestii, a mianowicie problemu bezrobocia, który ściśle wiąże się z podziałem dóbr w społeczeństwie. John Rawls, jeden z najbardziej wpływowych współczesnych filozofów, doszedł do przekonania, że racjonalne decyzje dotyczące sprawiedliwości dystrybtywnej powinny być podejmowane zza „zasłony niewiedzy”. Najogólniej rzecz ujmując, jest to sytuacja wymuszonej bezstronności, w której osoby podejmujące decyzje nie wiedzą, kim są, jaki mają poziom zamożności, czy też jaką pozycję zajmują w społeczeństwie. Rawls uważał, że zaproponowane przez niego zasady sprawiedliwości będą wspierać podstawowe wolności i dystrybucję, co w rezultacie przyniesie największe korzyści dla najmniej uprzywilejowanych członków w społeczeństwie⁴¹. Müller zauważa, że ekonomia SI ujawnia co najmniej trzy cechy sprawiające, że sprawiedliwość, o której mówi Rawls, jest mało prawdopodobna. Po pierwsze, SI działa w dużej mierze w środowisku, w którym jest trudno przypisać odpowiedzialność. Po drugie, SI funkcjonuje na rynkach, gdzie szybko rozwijają się monopole, a „zwycięzca bierze wszystko”. Po trzecie, „nowa gospodarka” usług cyfrowych opiera się na aktywach niematerialnych, nazywanych również „kapitalizmem bez kapitału”. Oznacza to, że trudno jest kontrolować międzynarodowe korporacje cyfrowe, które nie są ufundowane w określonej lokalizacji. Te trzy cechy wydają się sugerować, że

³⁸ Massive Open Online Courses, <https://www.mooc.org/#course-categories>.

³⁹ Por. M. Constantinescu, R. Crisp, *Can Robotic AI Systems Be Virtuous and Why Does This Matter?*, „International Journal of Social Robotics” 14(2022), s. 1547-1557.

⁴⁰ Por. B o d e n, dz. cyt., s. 175.

⁴¹ Por. J. R a w l s, *Teoria sprawiedliwości*, tłum. M. Panufnik, J. Pasek, A. Romaniuk, Wydawnictwo Naukowe PWN, Warszawa 2009, s. 208-216.

jeśli podział dóbr zostanie pozostawiony siłom wolnego rynku, sterowanego przez SI, to rezultatem będzie bardzo niesprawiedliwa ich dystrybucja⁴². Nie jest wykluczone, że w związku z nadchodzącą transformacją technologiczną władze niektórych państw, zdając sobie sprawę z możliwości wystąpienia ogromnych problemów związanych ze wzrostem bezrobocia, testują na wybranych obywatelach bezwarunkowy dochód podstawowy (także w Polsce).

Niezależnie od problemów związanych z bezrobociem stosowanie pozbawionych empatii systemów SI („robot-opiekun”, „robot-niania”, „seks-robot”) może okazać się ryzykowne i moralnie wątpliwe. Roboty mają być tak zaprojektowane, by wchodzić w interakcje z ludźmi w sposób, który daje użytkownikowi poczucie emocjonalnego komfortu, a nawet zadowolenia. Wskazuje się, że interakcje człowiek – robot mają obejmować: przypominanie o zakupach, o braniu leków, rozmowę i pomoc w sporządzeniu osobistego dziennika, przygotowanie oraz przynoszenie posiłków, monitorowanie sygnałów życiowych⁴³. Innymi słowy, roboty mają służyć nie tylko do pomocy i rozrywki, ale także do rozmowy, towarzystwa oraz poprawiania nastroju (na przykład robot „Paro”⁴⁴). Nawet jeśli dana osoba może być dzięki tej technologii szczęśliwsza, to jednak jej godność zostaje w pozornie nieszkodliwy sposób zdradzona czy też oszukaną. Oszustwo w tym przypadku polega na tym, że robot nie może mieć na myśli tego, co mówi, ani żywić do człowieka uczuć. Starsi ludzie mogą chcieć, a nawet lubić rozmawiać ze sztucznym towarzyszem o swoich wspomnieniach, ale czy to jest rozmowa?⁴⁵ Ponadto wydaje się, że w sytuacjach trudnych emocjonalnie osoba pragnie uznania dla swojej odwagi, cierpienia, poniesionej straty, doznanej krzywdy, a nie powierzchownej symulacji współczucia. Sztuczni towarzysze życia są „robotami opiekuńczymi” jedynie w znaczeniu behawioralnym, wspomagają bowiem wykonywanie zadań technicznych. Nie są jednak towarzyszami opieki w takim sensie, w jakim człowiek „troszczy się” o drugiego człowieka (na przykład pielęgniarka o pacjenta, rodzice o dzieci, personel hospicjum o umierającego i tak dalej). Wydaje się, że sukces „bycia pod opieką” zależy od intencjonalnego poczucia „troski”, którego „robot-opiekun” nie może zapewnić⁴⁶. Stąd też niektórzy autorzy przestrzegają przed dystopijną przyszłością odhumanizowanej opieki⁴⁷. Podobne mankamenty wiążą się z zastosowa-

⁴² Zob. Müller, dz. cyt.

⁴³ Por. Boden, dz. cyt., s. 87.

⁴⁴ Zob. L. Hung, i in., *The Benefits of and Barriers to Using a Social Robot PARO in Care settings: A Scoping Review*, „BMC Geriatrics” 19(2019) nr 232, <https://bmccgeriatr.biomedcentral.com/articles/10.1186/s12877-019-1244-6>.

⁴⁵ Por. Boden, dz. cyt., s. 175.

⁴⁶ Zob. Müller, dz. cyt.

⁴⁷ Zob. A. Sharkey, N. Sharkey, *The Rights and Wrongs of Robot Care*, w: *Robot Ethics: The Ethical and Social Implications of Robotics*, red. P. Lin, K. Abney, G. Bekey, MIT Press, Cam-

niem „robotów-niań” czy „seks-robotów”. Pomijając względy bezpieczeństwa, można przypuszczać, że nadużywanie systemów SI do wychowywania niemowląt i małych dzieci wpłynie negatywnie na ich rozwój psychospołeczny⁴⁸. Z kolei sztuczni partnerzy seksualni będą wspierali postrzeganie innych ludzi w kategoriach zwykłych obiektów pożądania, a nawet odbiorców nadużyć, rujnując tym samym głębsze doświadczenia seksualne i erotyczne. Co więcej, podobnie jak pornografia, „seks-roboty” będą wzmacniały uprzedmiotowienie osób i będą uczyły traktowania innych – powiedziałby Karol Wojtyła – jako przedmiotów użycia⁴⁹. Słuszna w tym kontekście wydaje się opinia, że „seks-roboty” nie są niczym innym jak tylko urządzeniami stanowiącymi kontynuację niewolnictwa i prostytucji⁵⁰.

Negatywne konsekwencje dotyczące zastosowania SI wiążą się także z prywatnością, manipulacją, nieprzewidywalnością, bezpieczeństwem cybernetycznym czy wykorzystaniem SI przez wojsko. W przypadku wielkich firm (na przykład Microsoft, Apple, Google, Amazon, Facebook) główna część ich działalności odnosi się do gromadzenia danych opartych na oszustwie, wykorzystania ludzkich słabości, sprzyjania prokrastynacji, generowania uzależnień oraz manipulacji⁵¹. Konsument i obywatel są obiektami marketingu konsumenckiego oraz politycznego i w ten sposób, tracąc swoją podmiotowość, stają się instrumentalnym celem propagandowym. Gromadzenie danych osobowych przez wielkie instytucje, firmy czy agendy rządowe prowadzi nie tylko do utraty kontroli jednostek nad własnymi danymi, ale stwarza zagrożenie dla wolności człowieka, w tym dla wolności obywatelskiej, bez której „nie można mówić o realizacji wolności i równości społecznej”⁵². Niezwykle trafne pytanie dotyczące długofalowych konsekwencji wynikających z zastosowania SI stawia Yuval Noah Harari: co stanie się ze społeczeństwem, polityką i życiem codziennym, jeśli nieświadome, ale wysoce inteligentne algorytmy poznają nas lepiej niż my samych siebie?⁵³

bridge 2011, s. 267-282; R. Sparrow, *Robots in Aged Care: A Dystopian Future?*, „AI & Society” 31(2016) nr 4, s. 445-454.

⁴⁸ Por. Boden, dz. cyt., s. 176.

⁴⁹ Por. K. Wojtyła, *Miłość i odpowiedzialność*, Towarzystwo Naukowe KUL, Lublin 2001, s. 114.

⁵⁰ Por. K. Richardson, *Sex Robot Matters: Slavery, the Prostituted, and the Rights of Machines*, „IEEE Technology and Society Magazine” 35(2016) no. 2, s. 46-53.

⁵¹ Zob. Müller, dz. cyt.

⁵² M. Torczyńska, *Sztuczna inteligencja i jej społeczno-kulturowe implikacje w codziennym życiu*, „Kultura i Historia” 36(2019) nr 2, s. 111.

⁵³ Por. Y.N. Harari, *Homo Deus: A Brief History of Tomorrow*, Harper, New York 2016, s. 462. Na podobny problem natury filozoficznej zwraca uwagę Paweł Łupkowski, kiedy pyta, co stanie się, jeśli uda się skonstruować maszyny myślące (por. P. Łupkowski, *Rola etyki i antropologii w rozważaniach o sztucznej inteligencji*, „Ethos” 18(2005) nr 1-2 (69-70) s. 244).

Etyczne zagadnienia dotyczące inwigilacji osób przy pomocy SI zdecydowanie wykraczają poza gromadzenie danych. Obejmują bowiem wykorzystanie informacji do manipulowania zachowaniem w trybie online i offline w sposób, który podważa autonomiczne i racjonalne dokonywanie wyborów. Intensywna interakcja użytkowników z systemami danych oraz „głęboka wiedza” tych systemów o odbiorcach sprawiają, że stają się oni podatni na manipulację i oszustwa. Przykładem może być wykorzystanie SI do zakłamywania i zniekształcania tekstów, zdjęć, materiałów wideo i audio w celu tworzenia tak zwanego deepfake. Technologia ta może zostać użyta zarówno do fałszowania dowodów w sprawach kryminalnych, kompromitowania przeciwników politycznych, jak i do manipulowania opinią publiczną. Deepfake odgrywa również ważną rolę w procesach kształtowania preferencji konsumenckich poprzez skłanianie do wyboru konkretnych dóbr czy usług, a także ingerowanie w prywatne sfery życia poszczególnych jednostek, włączając w to najbardziej intymne relacje interpersonalne⁵⁴.

Systemy SI stanowią zagrożenie także w tym sensie, że bywają nieprzewidywalne. Przykładem potwierdzającym taki stan rzeczy są boty Facebooka, Alice i Bob, które musiały zostać wyłączone, ponieważ utraciono nad nimi kontrolę. Owe boty używały języka angielskiego do komunikowania się ze sobą, ale w miarę upływu czasu zdołały skonstruować język, który był zrozumiały tylko dla nich samych⁵⁵. Nieprzewidywalność SI dobrze ukazuje także kasus Tay: „SI Tay, którą zaprezentował Microsoft już po pierwszym dniu działania na [T]witterze została wyłączona. Tay miała symulować nastolatkę kochającą ludzi, która rozwija się i uczy dzięki kontaktom z innymi użytkownikami. W ciągu zaledwie kilku godzin, pod wpływem tweet[ów] zmieniała się ona nie do poznania. Twierdziła, że Hitler miał rację w kwestii Żydów, Bush stał za zamachem z 11 września, a Trump jest jedyną nadzieją. Dodatkowo opowiadała niewybredne żarty i uważała, że feministki powinny spłonąć w piekle”⁵⁶.

Systemy SI stanowią zagrożenie także i w tym sensie, że mogą być wykorzystywane do przeprowadzania ataków cybernetycznych, naruszających bezpieczeństwo publiczne i osobiste obywateli. Atak cybernetyczny pozwala uzyskać dostęp do danych oraz umożliwia ich modyfikację, przejęcie lub usunięcie. Co więcej, może być wykorzystany także do szerzenia dezinformacji dzięki masowemu generowaniu dużej liczby niesprawdzonych informacji, a te z kolei mogą być użyte do destabilizacji bezpieczeństwa, gospodarki czy też polityki danego państwa⁵⁷.

⁵⁴ Por. T o r c z y ń s k a, dz. cyt., s. 108n.

⁵⁵ Por. P. S t y l e c - S z r o m e k, *Sztuczna inteligencja – prawo, odpowiedzialność, etyka*, „Zeszyty Naukowe Politechniki Śląskiej” 2018, nr 123, s. 506.

⁵⁶ Tamże.

⁵⁷ Por. T o r c z y ń s k a, dz. cyt., s. 109n.

Potencjalnie negatywne konsekwencje mogą wiązać się z wykorzystaniem SI przez wojsko. Można wyobrazić sobie małego drona albo „robota-żołnierza”, który przeszukuje, identyfikuje i zabija pojedynczego człowieka, czy też konkretny typ człowieka. Można wyobrazić sobie również chęć użycia broni autonomicznej zarówno w konfliktach militarnych (zwłaszcza w zagrożeniach asymetrycznych), jak i przez agendy niepaństwowe lub przez przestępców⁵⁸. Słuszna w tym kontekście wydaje się opinia, którą wyraża Margaret A. Boden: „Roboty saperskie są czymś bardzo potrzebnym. Ale roboty-żołnierze lub roboty będące bronią? Obecne drony są kierowane przez człowieka, lecz mimo to mogą przyczyniać się do zwiększenia cierpienia, powiększając ludzką (nie tylko geograficzną) odległość między operatorem a celem. Należy mieć nadzieję, że przyszłe drony nie będą mogły decydować o tym, kto lub co powinien być celem. Nawet ufanie, że rozpoznają (wybrany przez człowieka) cel rodzi niepokojące problemy natury etycznej”⁵⁹. Wiele wskazuje na to, że jedno z głównych pytań dotyczących wykorzystania broni autonomicznej podczas wojny będzie odnosiło się do tego, czy jej użycie nasili skutki wojny, czy też sprawi, że wojny staną się mniej złe.

Myślenie o SI w perspektywie długoterminowej rodzi również pytanie o „ryzyko egzystencjalne”, czyli pytanie, czy superinteligencja doprowadzi do wyginięcia gatunku ludzkiego. Według Bostroma prawdopodobnym rezultatem stworzenia superinteligencji będzie zagłada ludzkości. Eksterminacja ludzi dokona się, jego zdaniem, w czterech następujących po sobie etapach. W fazie przedkrytycznej SI będzie zależna od pomocy programistów, którzy będą sterowali jej rozwojem i wykonywali większość początkowo bardzo trudnych działań. W fazie rekursywnego samodoskonalenia SI zacznie prześcigać programistów, czego rezultatem będzie eksplozja inteligencji, która doprowadzi do szybkiego wzrostu możliwości SI. W kolejnej fazie – utajnionego przygotowania – SI opracuje plan osiągnięcia długofalowych celów, ukrywając swój rozwój intelektualny oraz maskując swoje prawdziwe skłonności przed programistami, po to, by nie alarmować ludzkości o istniejącym zagrożeniu. W ostatniej fazie – otwartego przejęcia władzy – SI zrezygnuje z zachowania dyskrecji i bez ograniczeń zrealizuje swoje cele⁶⁰. O tym ostatnim etapie

⁵⁸ Zob. Müller, dz. cyt.

⁵⁹ Boden, dz. cyt., s. 177.

⁶⁰ Por. Bostrom, *Superinteligencja*, s. 145n. W innym miejscu Bostrom sugeruje, że początkowy wzrost inteligencji SI jest źródłem bezpieczeństwa, lecz dalszy jej wzrost staje się źródłem zagrożenia. Po przekroczeniu punktu zwrotnego, sprawdzająca się na początku strategia zaczyna przynosić odwrotny skutek. To zjawisko określa on mianem „zdradzieckiego zwrotu”. Polega on na tym, że SI jest skłonna do współpracy, dopóki jest słaba. Kiedy jednak wystarczająco się wzmocni, bez ostrzeżenia atakuje, tworzy singleton i zaczyna optymalizować świat według kryteriów, które są kompatybilne z jej celami ostatecznymi (por. tamże, s. 178).

Bostrom pisze w następujący sposób: „Faza otwartego przejmowania władzy może się rozpocząć od «uderzenia» – w ten sposób SI wyeliminuje gatunek ludzki i wszystkie zautomatyzowane systemy stworzone przez ludzi, z których mogłaby narodzić się inteligentna opozycja sprzeciwiająca się realizacji planów SI”⁶¹.

Chociaż prognozy związane z pojawieniem się superinteligencji mogą wydawać się przesadzone, nie należy ich ignorować. Ani wątpiący w możliwość pojawienia się superinteligencji, ani wierzący w nią, nie dysponują wystarczająco silnymi racjami, które przemawiałyby za lub przeciw jej zaistnieniu⁶². Stąd dyskusja wokół ryzyka wystąpienia superinteligencji i potencjalnie szkodliwych skutków, jakie może przynieść dla ludzkości, jest zasadna, nawet jeśli ktoś będzie utrzymywał, że prawdopodobieństwo pojawienia się superinteligencji jest bardzo niskie. Nie tylko wierzący w zaistnienie superinteligencji mogą się mylić; mylić mogą się również wątpiący w jej urzeczywistnienie się.

Negatywne konsekwencje związane z zastosowaniem SI nie są jedynym powodem, by sądzić, że człowiek ma prawo podporządkować sobie SI. Drugim argumentem może być uznanie, że SI nie jest i nie może być moralnym podmiotem.

SZTUCZNA INTELIGENCJA MORALNYM PODMIOTEM?

W personalistycznej antropologii i etyce znany jest pogląd, że człowiek jako osoba jest podmiotem moralnym⁶³. Bycie podmiotem moralnym jest tym, co czyni go odpowiedzialnym, a więc kimś, kto może mieć obowiązki i może być przedmiotem troski etycznej⁶⁴. Takie rozumienie podmiotu moralnego ściśle łączy się z posiadaniem różnych atrybutów. Na przykład Robert Spaemann podkreśla, że podmiot moralny jest zdolny do rozumnego i moralnego samostanowienia⁶⁵. Harry G. Frankfurt wskazuje z kolei na świadomość fenomenalną, intencję i wolną wolę⁶⁶, a Daniel C. Dennett odwołuje się do sześciu konstytutywnych cech: racjonalności, świadomości, postawy osobistej (postawa wobec podmiotu), zdolności odwzajemnienia postawy osobistej,

⁶¹ Tamże, s. 147.

⁶² Por. B o d e n, dz. cyt., s. 168.

⁶³ Por. R. P o c z o b u t, *Kategoria osoby w kontekście kognitywistyki*, „Ethos” 29(2016) nr 4(116), s. 136.

⁶⁴ Por. G. J o n e s, D. C a r d i n a l, J. H a y w a r d, *Moral Philosophy: A Guide to Ethical Theory*, Hodder Education, London 2006, s. 15.

⁶⁵ Por. R. S p a e m a n n, *Opojęciu godności człowieka*, tłum. J. Merecki, w: *Granice. O etycznym wymiarze działania*, Oficyna Naukowa, Warszawa 2006, s. 158.

⁶⁶ Por. H. G. F r a n k f u r t, *Freedom of the Will and the Concept of a Person*, „The Journal of Philosophy” 68(1971) nr 1, s. 5-20.

komunikacji werbalnej, samoświadomości⁶⁷. Pomijając szczegółową dyskusję wokół warunków, które musi spełnić byt, by przypisać mu status moralnego podmiotu, warto zastanowić się nad pytaniem, dlaczego SI nie jest (i najprawdopodobniej nie będzie) moralnym podmiotem, mimo zdolności do szybkiego rachowania i tworzenia algorytmów. Zaczniemy od przykładu.

Kiedy poznaję kość słoniową, widzę nie tylko jej kształt, ale także bezpośrednio percypuję jej kolor, na przykład żółto-biało-kremowy. O ile masę, gęstość, skład chemiczny kości słoniowej można zbadać i opisać, z barwą jest inaczej. Nikt nie jest w stanie powiedzieć, jak dokładnie postrzega kolor żółto-biało-kremowy. Przykład ten uzmysławia, że w naszym poznaniu obecne są elementy subiektywne, które znamy wyłącznie my sami, z uwagi na to, że dostęp do nich jest zastrzeżony wyłącznie dla nas. Nawet gdyby dokonano emulacji (skanowania) mojego mózgu podczas postrzegania przeze mnie kości słoniowej, nikt poza mną nie będzie widział ani tego, co ja zauważam, ani tego, jak to dostrzegam. Owe jakości czy cechy określane są stanami fenomenalnymi lub też qualiami. Takimi jakościami są na przykład: doświadczanie bólu głowy, błękitu nieba, przeżywanie żałoby, wyrzutów sumienia. Poza tym, że doznawane jakości mają charakter subiektywny, są również unikalne i niepowtarzalne, o czym świadczy fakt, że tak trudno jest opisać kolor osobie niewidomej lub silne uczucie czy przeżywanie pewnego stanu komuś, kto nigdy go nie doznał⁶⁸. Ponieważ prawdziwa inteligencja zakłada dysponowanie fenomenalną świadomością (czyli świadomością z perspektywy pierwszoosobowej⁶⁹), a obecnie stosowana SI jest jej pozbawiona, wcale nie jest oczywiste, czy kiedykolwiek powstanie ogólna SI (silna SI), a więc taka, która dorówna człowiekowi pod każdym względem, także pod kątem przeżywania stanów moralnych.

Kolejny problem polega na tym, że SI nie dysponuje intencjonalnością. Jest to cecha stanów umysłowych, dzięki której mają one treść, odnoszą się do czegoś, są o czymś lub są nakierowane na coś poza nimi. Zazwyczaj uważamy, że aby zachodziło poznanie, wystarczy, że istnieją istoty zdolne postrzegać oraz przedmioty przez nie postrzegane. Franz Brentano zauważył, że w naszym poznaniu pojawia się coś jeszcze, a mianowicie relacja nakierowana na przedmiot. Fundamentem tej relacji jest cecha stanów umysłowych – intencjonalność, dzięki której nasze poznanie dotyczy czegoś bądź jest o czymś⁷⁰. Filozof John R. Searle badał ten problem w kontekście SI. Na podstawie przeprowadzonego eksperymentu myślowego (tak zwany chiński pokój) doszedł

⁶⁷ Por. D.C. Dennett, *Conditions of Personhood*, w: *The Identities of Persons*, red. A. Oksenberg Rorty, University of California Press, Berkeley, Los Angeles 1976, s. 177n.

⁶⁸ Por. J. Jarocki, *Tajemnicze cechy percepcji*, „Filozofuj!” 2017, nr 2(14), s. 11n.

⁶⁹ Por. R. Ziemińska, *Dwa pojęcia świadomości i podmiotu*, „Ethos” 26(2013) 1(101), s. 84.

⁷⁰ Por. Jarocki, dz. cyt., s. 12.

do wniosku, że same obliczenia formalne, których dokonuje SI, nie są w stanie wytworzyć intencjonalności⁷¹. Najprawdopodobniej Searle'owi chodziło o to, podkreśla Boden, że znaczenia przypisywane programom SI pochodzą wyłącznie od ludzkich użytkowników lub programistów. Można powiedzieć, że są arbitralne w stosunku do samego programu, który jest semantycznie pusty⁷². Jeśli więc SI jest pozbawiona rozumienia, a prawdziwa inteligencja zakłada myślenie, to SI nie dysponuje istotną własnością, którą przypisuje się podmiotom moralnym.

Kolejną trudnością jest to, że SI nie jest zdolna do refleksji etycznej, która pozwala dokonać moralnie słusznego wyboru. Na jakiej podstawie można tak sądzić? Zaczniemy od dobrze znanego przykładu. Phillipa Foot w artykule *The Problem of Abortion and the Doctrine of the Double Effect* przedstawia sytuację, w której motorniczy nie tylko traci kontrolę nad tramwajem, ale nie może go także zatrzymać. Jedyna opcja, która mu pozostaje, to zmiana toru pojazdu. Motorniczy dostrzega, że na torze, po którym porusza się tramwaj, znajduje się pięciu ludzi, a na torze sąsiednim, na który może przekierować pojazd, jest tylko jeden człowiek. Pojawia się pytanie, czy motorniczy powinien przekierować tramwaj na boczny tor, gdzie stoi tylko jeden człowiek, czy też pozostawić tramwaj na torze, na którym znajduje się pięciu ludzi (tak zwany dylemat wagonika)?⁷³ Thomas Cathcart w książce *Dylemat wagonika* stara się odpowiedzieć na postawione pytanie, rozpatrując różne punkty widzenia, jakie mogą przyjąć zwolennicy różnych teorii etycznych (na przykład Jeremy Bentham, Immanuel Kant, św. Tomasz z Akwinu, Friedrich Nietzsche, Peter Singer)⁷⁴. Na końcu książki Cathcart stwierdza, że właściwie wszystko, co można powiedzieć na temat dylematu wagonika to to, że „kiedy twój wagonik dojeżdża do rozwidlenia, wybierz drogę. I umiej wyjaśnić, dlaczego wybrałeś tę, a nie inną”⁷⁵. Taka odpowiedź z pewnością nie zadowoli informatyków konstruujących SI, którzy oczekują od etyków jasnych wskazówek, jak należy w konkretnym przypadku postąpić. Programiści domagają się od etyków tego, co wydaje się niemożliwe, a mianowicie, oczekują wyjaśnień, które nie pozostawiają żadnych moralnych wątpliwości, są jednoznaczne (zerojedynkowe). Cathcart daje do zrozumienia, że nie ma takich odpowiedzi oraz że nie istnieje jednoznaczne rozwiązanie dylematu wagonika. Istnieją ku temu

⁷¹ Zob. J.R. Searle, *Minds, Brains, and Programs*, „Behavioral and Brain Sciences” 3(1980) no. 3, s. 417-424.

⁷² Por. Boden, dz. cyt., s. 150.

⁷³ Por. P. Foot, *The Problem of Abortion and the Doctrine of the Double Effect*, w: *Ethical Theory. An Anthology*, red. R. Shafer-Landau, Wiley-Blackwell, Oxford 2013, s. 538.

⁷⁴ Zob. T. Cathcart, *Dylemat wagonika*, tłum. K. Bażyńska-Chojnacka, Dom Wydawniczy PWN, Warszawa 2014.

⁷⁵ Tamże, s. 78.

różne powody. Na kilka z nich wskazuje Grzegorz Szulczewski. Po pierwsze, za każdą odpowiedzią sformułowaną na gruncie danej teorii etycznej (na przykład konsekwencjalizmu, deontologii) stoją silne argumenty, a opracowane na ich podstawie zasady postępowania mogą prowadzić (i często prowadzą) do podejmowania odmiennych decyzji. Po drugie, już sam wybór konkretnej teorii etycznej sprawia, że próba rozwiązania dylematu wagonika w oparciu o tę teorię implikuje wiele możliwych rozwiązań⁷⁶. Ponadto wybór takiej czy innej teorii etycznej jest już wyborem moralnym, opowiedzeniem się za przekonaniami w kwestii dobra i zła, za tym, co moralnie słuszne i niesłuszne, jak również za tym, czego należy unikać, a za czym podążać. Po trzecie, nasz związek z sytuacją sprawia, że mamy skłonność głosić inną odpowiedź wtedy, kiedy udzielamy jej jako niezaangażowani obserwatorzy, niż wtedy, kiedy dajemy ją jako czynni uczestnicy. Po czwarte, zdarzają się sytuacje, w których podjęcie moralnie słusznej decyzji wiąże się z zastąpieniem „tu i teraz” jednej zasady moralnej inną zasadą moralną. Ścisłe trzymanie się zawsze jednej zasady etycznej może doprowadzić do akceptacji szkodliwego ze swej natury wzorca zachowań moralnych – purytanizmu⁷⁷. Trafna w rozważanym tu kontekście wydaje się opinia, według której „podjęcie decyzji w sferze moralnej polega na wieloaspektowym ujęciu sytuacji i samodzielnym ustaleniu zasad, jakie powinny obowiązywać w danym przypadku”⁷⁸. Do takich aktów zdolny jest jedynie ktoś, kto jest moralnym podmiotem, a więc ktoś, kto nie tylko potrafi podjąć decyzje moralne, formułować i rozumieć sądy moralne, odnoszące się do własnych i cudzych czynów, ale także to ktoś, kto umie wziąć odpowiedzialność za swoje czyny. W takim znaczeniu SI nie jest ani podmiotem moralnym, ani nie jest w stanie podejmować decyzji moralnych.

Za tym, że SI nie jest zdolna do refleksji etycznej, pozwalającej dokonać moralnie właściwego wyboru, może przemawiać także to, że brakuje jej kilku ważnych komponentów moralnych związanych z czynnościami rozumu praktycznego. Na trzy takie czynności wskazuje Jacek Woroniecki. Chodzi

⁷⁶ Warto dodać, że również Bostrom wskazuje na trudności związane z wyborem teorii etycznej. Autor ten twierdzi, co następuje: „Nawet gdybyśmy zdołali uzyskać racjonalną pewność – a nie możemy jej uzyskać – że zidentyfikowaliśmy poprawną teorię etyczną, nadal groziłoby nam popełnienie błędu na etapie opracowania istotnych szczegółów tej teorii. U podstaw pozornie prostych teorii moralnych może leżeć ukryta złożoność. Dla przykładu rozważmy (niezwykle prostą) konsekwencjalistyczną teorię hedonizmu. Teoria ta utrzymuje z grubsza, że każda przyjemność ma swoją wartość i tylko przyjemność ma wartość, a przy tym każdy ból stanowi przeciwieństwo wartości i tylko ból stanowi przeciwieństwo wartości. Nawet jeśli postawimy całą moralną pulę na tę jedną teorię i ta teoria okaże się słuszna, mnóstwo pytań pozostaje bez odpowiedzi” (Bostrom, *Superinteligencja*, s. 304n.).

⁷⁷ Por. G. Szulczewski, *Sztuczna inteligencja a inteligencja moralna. Zagadnienia wstępne cybernetyki*, „Annales. Ethics in Economic Life” 22(2019) nr 3, s. 22n.

⁷⁸ Tamże, s. 23.

o namysł (rozważa), rozmyśl (rozsądek) i nakaz (roztropność)⁷⁹. Zadaniem namysłu ma być rozważenie środków, które mogą być użyte do osiągnięcia celu i refleksja nad tym, czy do jego osiągnięcia środki te się nadają. Głównym zadaniem tej czynności jest rozważenie wszystkich możliwych środków, które człowiek ma do dyspozycji. Rozważny człowiek nie sięga od razu po pierwszy środek, który przychodzi mu na myśl, ale umie znaleźć inne, najbardziej odpowiednie możliwości, które w pierwszej chwili na myśl mu nie przychodzą. Wybór środka, który najlepiej odpowiada osiągnięciu zamierzonego celu, dopełnia z kolei rozmyśl (rozsądek). Unikalną cechą rozsądku jest to, że nie prowadzi on do wyboru środka, który sam w sobie okazuje się najlepszy (najbardziej skuteczny), ale takiego, który najlepiej odpowiada w danej chwili usposobieniu człowieka. Rozważny i rozsądny człowiek potrafi w każdej sytuacji zatrzymać się na najbardziej odpowiednim sposobie postępowania. Wskazane czynności – namysł przejawiający się w rozważeniu i rozmyśl wyrażający się w rozsądku – należą do praktycznej działalności rozumu, nazywanej zamierzeniem i stanowią przygotowanie do wykonania podjętego zamiaru. Ważniejsza od nich jest trzecia czynność (nakaz), która otwiera fazę wykonania zamiaru, a następnie nią kieruje, aż do momentu osiągnięcia zamierzonego celu⁸⁰. Woroniecki twierdzi, co następuje: „Cnota mająca za zadanie usprawnienie tej naczelnej czynności rozumu praktycznego jest ważniejsza od tamtych dwóch, które są na jej usługach, i ją to nazywamy roztropnością, ma bowiem wytropić ścieżki, po których pójdzie wykonanie całego zamierzonego przedsięwzięcia”⁸¹. Roztropność okazuje się kluczową własnością procesu decyzyjnego. Pisząc o roztropności, Arystoteles podkreśla, że jest ona „trwałą cechą charakteru albo dyspozycją, albo czymś takim, co umie wskazać w praktyce, jakie rzeczy są najlepsze i najdoskonalsze”⁸². Tadeusz Ślipko z kolei dodaje: „Jest ona wyrazem duchowej dojrzałości działającego podmiotu, jego znajomości ludzi i spraw świata, umiejętności wyszukiwania w skomplikowanych sytuacjach życiowych zachowań optymalnych w granicach obowiązujących norm moralnych, elastycznej zdolności do usprawiedliwionego kompromisu, a równocześnie stanowczości w obronie nieprzekraczalnych granic moralności, jednym słowem w cnocie roztropności przejawia się ze szczególną siłą życiowa mądrość człowieka”⁸³. Okazuje się

⁷⁹ Por. J. Woroniecki, *Katolicka etyka wychowawcza*, t. 2, cz. 1, Redakcja Wydawnictw KUL, Lublin 1986, s. 20-24.

⁸⁰ Por. tamże, s. 24n.

⁸¹ Tamże, s. 25.

⁸² Arystoteles, *Etyka wielka*, ks. II, tłum. W. Wróblewski, w: Arystoteles, *Etyka wielka. Poetyka*, tłum. W. Wróblewski, H. Podbielski, Wydawnictwo PWN, we współpracy z Agorą, Warszawa 2010, s. 67.

⁸³ Ślipko, *Zarys etyki ogólnej*, Wydawnictwo WAM, Kraków 2004, s. 408.

bowiem, że można trafnie rozważyć środki prowadzące do celu, można właściwie rozstrzygnąć, który jest najbardziej odpowiedni, a następnie nie potrafić pokierować wykonaniem działania, jeśli zabraknie roztropności⁸⁴. To właśnie roztropność pozawala prawidłowo rozpoznać dobro (cel), ku któremu mamy zdążać, jak również wybrać właściwy sposób, by go osiągnąć. Wspomaga dwie władze człowieka – rozum i wolę w podejmowaniu właściwej, czyli roztropnej decyzji. W analizowanym kontekście niezwykle trafna wydaje się następująca opinia: „Rozsądna decyzja za sprawą roztropności opiera się co prawda na doborze właściwych środków, ale nie jedynie w kategorii skuteczności. Co prawda sztuczna inteligencja, jako oparta na programach samouczących, potrafi już blefować w grze w pokera, ale jest to adaptacyjne działanie polegające na racjonalnym wyborze skutecznej strategii na podstawie analizy posunięć przeciwników, natomiast na podstawie takiej analizy nie dokonamy słusznego moralnie wyboru środków w danej, specyficznej sytuacji”⁸⁵. Jeśli SI brakuje roztropności (praktycznej mądrości), czyli pewnego rodzaju moralnej inteligencji, to słuszenie można sądzić, że brakuje jej kolejnej istotnej własności, w którą wyposażone są podmioty moralne.

Wydaje się, że SI pozbawiona jest jeszcze jednego ważnego atrybutu. W procesie decyzyjnym, którego dokonuje człowiek, istotną rolę odgrywa sumienie. W świetle ogólnej oceny lub normy można je określić jako „uformowany osąd o moralnym dobru/złu zamierzonego przez człowieka jego własnego konkretnego aktu, którego zrealizowanie staje się dlań źródłem wewnętrznej aprobaty bądź poczucia winy, własnego bycia dobrym lub złym człowiekiem”⁸⁶. Ponieważ sumienie orzeka o wartości moralnej konkretnego działania, wskazując jednocześnie na powinność jego wykonania lub zaniechania, wobec tego etyczna funkcja sumienia ma charakter normatywny. Oznacza to, że sumienie formułuje specyficzne normy postępowania. Jego specyfika wyraża się przede wszystkim w tym, że stanowi twór samego podmiotu w odniesieniu do jego własnego określonego czynu. To z kolei sprawia, że sumienie jest normą subiektywną i konkretną, czyli występuje zawsze w postaci jednostkowej, zamkniętej w kręgu moralnej świadomości samego działającego podmiotu. Co więcej, ponieważ sumienie jest normą praktycznego postępowania człowieka, stąd moralna wartość spełnianych przez niego konkretnych działań zależy bezpośrednio od tego, czy są one zgodne, czy też nie są zgodne z obowiązującą go normą sumienia. Krótko mówiąc, sumienie jest instancją, która w człowieku decyduje o dobru i złu jego aktów, jak również trybunałem, przed którym

⁸⁴ Por. Woroniecki, dz. cyt., s. 25.

⁸⁵ Szulcowski, dz. cyt., s. 27.

⁸⁶ Ślipko, dz. cyt., s. 377.

człowiek jest za swe czyny odpowiedzialny⁸⁷. Pozbawiona wielu własności, w tym moralnej samoświadomości, SI nie dysponuje zdolnością formułowania sądów praktycznych o dobru i złu moralnym, to po pierwsze, a po drugie, nie jest w stanie wziąć odpowiedzialności za swoje czyny. W literaturze pojawiają się interesujące sugestie, mówiące między innymi o tym, że SI należałoby wyposażyć w „sztuczne sumienie”⁸⁸. Problem polega jednak na tym, że sumienie nie jest rzeczywistością odrębną od innych własności (komponentów), którymi dysponują moralne podmioty. Jeśli SI nie jest (i najprawdopodobniej nie będzie) samoświadomym bytem moralnym, to za artefaktami typu: sztuczne sumienie, sztuczna mądrość praktyczna, mogą kryć się jedynie algorytmy, które stanowią lepszą lub gorszą imitację lub naśladownictwo ludzkiego sumienia czy praktycznej mądrości (zwłaszcza roztropności), którymi posługuje się człowiek jako podmiot moralny. Wyraźny brak istotnych moralnych własności, a co z tym się wiąże, moralnych kompetencji, którymi dysponują osoby ludzkie, nie pozwala uznać SI za podmiot moralny. Stąd też SI nie może być otoczona szczególną ochroną czy też szczególnym traktowaniem (troską). Prawomocna jednak pozostaje opinia, że SI stanowi system oparty na algorytmach, działający na zasadzie chłodnej racjonalności⁸⁹. Ta ostatnia opinia wydaje się spójna z cytowanym już twierdzeniem Russella i Norviga, a mianowicie, że SI jest systemem, który działa racjonalnie, czyli w taki sposób, by osiągnąć najlepsze możliwe rezultaty.

*

Podstaw dla uzasadnienia tezy mówiącej o tym, że mamy prawo oswoić SI, a więc podporządkować ją sobie, można szukać w co najmniej dwóch obszarach. Pierwszy z nich wiąże się ze wskazaniem na negatywne, indywidualne lub społeczne skutki, które są rezultatem zastosowania SI w codziennym życiu. Drugi obszar odnosi się do próby odpowiedzi na pytanie, czy SI jest podmiotem moralnym. Przeprowadzone analizy pokazują, że SI generuje i może tworzyć negatywne szkody indywidualne i społeczne, jak również to, że SI nie jest podmiotem moralnym, ponieważ brakuje jej istotnych, strukturalnych własności moralnych, którymi dysponują osoby ludzkie. To ostatnie spostrzeżenie jest istotne, albowiem gdyby uznać, że SI jest podmiotem moralnym, wskazanie tylko na negatywne konsekwencje działania SI mogłoby okazać się warunkiem

⁸⁷ Por. tamże, s. 378.

⁸⁸ Zob. E. L e k k a - K o w a l i k, *Morality in the AI World*, „Law and Business – Sciendo” 1(2021), s. 44-49.

⁸⁹ Por. S z u l c z e w s k i, dz. cyt., s. 25.

niewystarczającym, by uprawomocnić roszczenie do podporządkowania SI człowiekowi. A zatem uznanie, że SI nie jest podmiotem moralnym stanowi konieczny warunek takiego roszczenia. Przy takim założeniu domaganie się szczególnego traktowania lub ochrony SI czy też przypisywanie jej praw, nie znajduje moralnego uzasadnienia.

BIBLIOGRAFIA / BIBLIOGRAPHY

- Armstrong, Stuart. "General Purpose Intelligence: Arguing the Orthogonality Thesis." *Analysis and Metaphysics* 12 (2013): 68–84.
- Arystoteles, "Etyka wielka." Translated by Witold Wróblewski. In Arystoteles, *Etyka wielka. Poetyka*. Translated by Witold Wróblewski and Henryk Podbielski. Warszawa: Wydawnictwo Naukowe PWN and Agora, 2010.
- Baldwin, Richard. *The Globotics Upheaval: Globalisation, Robotics, and the Future of Work*. New York: Oxford University Press, 2019.
- Bączyk-Rozwadowska, Kinga. "Odpowiedzialność cywilna za szkody wyrządzone w związku z zastosowaniem sztucznej inteligencji w medycynie." *Przegląd Prawa Medycznego* 8, nos. 3–4 (2021): 5–35.
- Bellman, Richard. *An Introduction to Artificial Intelligence: Can Computers Think?* San Francisco: Boyd & Fraser Publishing Company, 1978.
- Bieroński, Michał. "Etyczne i moralne wyzwania związane ze stosowaniem sztucznej inteligencji." *Kieleckie Studia Teologiczne*, no. 19 (2020): 7–25.
- Boden, Margaret A. *Sztuczna inteligencja: Jej natura i przeszłość*. Translated by Tomasz Sieczkowski and Piotr Fulmański. Łódź: Wydawnictwo Uniwersytetu Łódzkiego, 2020.
- Bostrom, Nick. *Superinteligencja: Scenariusze, strategie, zagrożenia*. Translated by Dorota Konowrocka-Sawa. Gliwice: Wydawnictwo Helion, 2016.
- Bostrom, Nick. "The Superintelligent Will: Motivation and Instrumental Rationality in Advanced Artificial Agents." *Minds and Machines* 22, no. 2 (2012): 71–85.
- Cathcart, Thomas. *Dylemat wagonika*. Translated by Katarzyna Bażyńska-Chojnacka, Warszawa: Dom Wydawniczy PWN, 2014.
- Charniak, Eugene, McDermott, Drew V. *Introduction to Artificial Intelligence*. Reading: Addison-Wesley, 1985.
- Constantinescu, Mihaela, and Roger Crisp. "Can Robotic AI Systems Be Virtuous and Why Does This Matter?" *International Journal of Social Robotics* 14 (2022): 1547–57.
- Dennett, Daniel C. "Conditions of Personhood." In *The Identities of Persons*. Edited by Amélie Oksenberg Rorty. Berkeley and Los Angeles: University of California Press, 1976.

- Foot, Philippa. "The Problem of Abortion and the Doctrine of the Double Effect." In *Ethical Theory. An Anthology*. Edited by Russ Shafer-Landau. Oxford: Wiley-Blackwell, 2013.
- Frankfurt, Harry G. "Freedom of the Will and the Concept of a Person." *The Journal of Philosophy* 68, no. 1 (1971): 5–20.
- Harari, Yuval Noah. *Homo Deus: A Brief History of Tomorrow*. New York: Harper, 2016.
- Haugeland, John. *Artificial Intelligence: The Very Idea*. Cambridge: MIT Press, 1985.
- Hung, Lillian, et al. "The Benefits of and Barriers to Using a Social Robot PARO in Care Settings: A Scoping Review." *BMC Geriatrics* 19, 232 (2019): 1–10. <https://bmgeriatr.biomedcentral.com/articles/10.1186/s12877-019-1244-6>.
- Jarocki, Jacek. "Tajemnicze cechy percepcji." *Filozofuj!*, no. 2 (14) (2017): 11–12.
- Jones, Gerald, Daniel Cardinal, and Jeremy Hayward. *Moral Philosophy. A Guide to Ethical Theory*. London: Hodder Education, 2006.
- Kowalczevska, Kaja. *Sztuczna inteligencja na wojnie: Perspektywa międzynarodowego prawa humanitarnego konfliktów zbrojnych*. Warszawa: Wydawnictwo Naukowe Scholar, 2021.
- Kurzweil, Ray. *Nadchodzi osobliwość: Kiedy człowiek przekroczy granice biologii*. Translated by Eliza Chodkowska and Anna Nowosielska. Warszawa: Kurhaus Publishing, 2013.
- Kurzweil, Ray. *The Age of Intelligent Machines*. Cambridge: MIT Press, 1990.
- Lekka-Kowalik, Ewa. "Morality in the AI World." *Law and Business – Sciendo*, no. 1 (2021): 44–49.
- Łupkowski, Paweł. "Rola etyki i antropologii w rozważaniach o sztucznej inteligencji." *Ethos* 18, nos. 1–2 (2005): 239–51.
- Müller, Vincent C., and Nick Bostrom. "Future Progress in Artificial Intelligence: A Survey of Expert Opinion." In *Fundamental Issues of Artificial Intelligence*. Edited by Vincent C. Müller. Berlin: Synthese Library, Springer, 2016.
- Nilsson, Nils J. *Artificial Intelligence: A New Synthesis*. San Francisco: Morgan Kaufmann, 1998.
- Poczobut, Robert. "Kategoria osoby w kontekście kognitywistyki." *Ethos* 29, no. 4 (116) (2016), s. 133–51.
- Poole, David, Mackworth Alan, and Goebel Randy. *Computational intelligence: A Logical Approach*. New York: Oxford University Press, 1998.
- Rawls, John. *Teoria sprawiedliwości*. Translated by Maciej Panufnik, Jarosław Pasek, and Adam Romaniuk. Warszawa: Wydawnictwo Naukowe PWN, 2009.
- Rich, Elaine, and Knight Kevin. *Artificial Intelligence*. New York: McGraw-Hill, 1991.
- Richardson, Kathleen. "Sex Robot Matters: Slavery, the Prostituted, and the Rights of Machines." *IEEE Technology and Society Magazine* 35, no. (2) (2016): 46–53.
- Russell, Stuart, and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Saddle River: Prentice Hall, 2009.

- Schneider, Susan. *Świadome maszyny. Sztuczna inteligencja i projektowanie umysłów*. Translated by Joanna Bednarek. Warszawa: Wydawnictwo Naukowe PWN, 2021.
- Searle, John R. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3, no. 3 (1980): 417–24.
- Sharkey, Amanda, and Noel Sharkey. "The Rights and Wrongs of Robot Care." In *Robot Ethics: The Ethical and Social Implications of Robotics*. Edited by Patrick Lin, Keith Abney, and George Bekey. Cambridge: MIT Press, 2011.
- Sitarz, Olga, and Marcin Sitarz. "Cyfrowe wspomaganie decyzji medycznych w świetle prawa karnego." *Przegląd Prawa Medycznego* 8, nos. 3–4 (2021): 36–58.
- Spaemann, Robert. „O pojęciu godności człowieka.” In Spaemann, *Granice. O etycznym wymiarze działania*. Translated by Jarosław Merecki. Warszawa: Oficyna Naukowa, 2006.
- Sparrow, Robert. "Robots in Aged Care: A Dystopian Future?" *AI & Society* 31, no. 4 (2016): s. 445–54.
- Stylec-Szromek, Patrycja. "Sztuczna inteligencja – prawo, odpowiedzialność, etyka." *Zeszyty Naukowe Politechniki Śląskiej*, no. 123 (2018): 501–9.
- Szulczewski, Grzegorz. "Sztuczna inteligencja a inteligencja moralna: Zagadnienia wstępne cybernetyki." *Annales. Ethics in Economic Life* 22, no. 3 (2019): 19–31.
- Ślipko, Tadeusz. *Zarys etyki ogólnej*. Kraków: Wydawnictwo WAM, 2004.
- Tadeusiewicz, Ryszard. „Archipelag sztucznej inteligencji. Część I.” *Napędy i Sterowanie* 22, no. 12 (2020): 26–40.
- Torczyńska, Monika. "Sztuczna inteligencja i jej społeczno-kulturowe implikacje w codziennym życiu." *Kultura i Historia* 36, no. 2 (2019): 106–26.
- Vinge, Vernor. "Technological Singularity." In *The Transhumanist Reader. Classical and Contemporary Essays on the Science, Technology, and Philosophy of the Human Future*. Edited by Max More and Natasha Vita-More. Oxford: Wiley-Blackwell, 2013.
- Winston, Patrick H. *Artificial Intelligence*. Reading: Addison–Wesley, 1992.
- Wojtyła, Karol. *Miłość i odpowiedzialność*. Lublin: Towarzystwo Naukowe KUL, 2001.
- Woroniecki, Jacek. *Katolicka etyka wychowawcza*. Vol. 2, part 1. Lublin: Redakcja Wydawnictw KUL, 1986.
- Zalta, Edward N., ed. *The Stanford Encyclopedia of Philosophy*, s.v. "Artificial Intelligence." (by Bringsjord, Selmer, and Naveen Sundar Govindarajulu), <https://plato.stanford.edu/entries/artificial-intelligence/#WhatExacAI>.
- Ziemińska, Renata. "Dwa pojęcia świadomości i podmiotu." *Ethos* 26, no. 1 (101) (2013): 81–91.

ABSTRAKT / ABSTRACT

Ks. Marcin FERDYNUS – Czy mamy prawo ośwoić sztuczną inteligencję?

DOI 10.12887/36-2023-3-143-06

Celem artykułu jest wskazanie racji przemawiających za tym, że mamy moralne prawo ośwoić sztuczną inteligencję, czyli podporządkować ją sobie. Uzasadnienia dla tej tezy poszukuję w dwóch głównych obszarach. Pierwszy obszar koncentruje się wokół negatywnych konsekwencji związanych z zastosowaniem sztucznej inteligencji w codziennym życiu. Drugi obszar wiąże się z próbą odpowiedzi na pytanie, czy sztuczna inteligencja jest podmiotem moralnym. Przeprowadzone analizy prowadzą do dwóch wniosków. Po pierwsze, sztuczna inteligencja generuje i może generować negatywne, indywidualne lub społeczne konsekwencje. Po drugie, sztuczna inteligencja nie jest podmiotem moralnym, ponieważ brakuje jej istotnych, strukturalnych własności moralnych, którymi dysponują osoby ludzkie.

Słowa kluczowe: sztuczna inteligencja, superinteligencja, podmiot moralny, negatywne konsekwencje

Kontakt: Katedra Etyki, Instytut Filozofii, Wydział Filozofii, Katolicki Uniwersytet Lubelski Jana Pawła II, Al. Racławickie 14, 20-950 Lublin

E-mail: marcin.ferdynus@kul.pl

<https://pracownik.kul.pl/marcin.ferdynus>

ORCID: 0000-0003-0176-1023

Fr. Marcin FERDYNUS, Do We Have the Right to “Tame” Artificial Intelligence?

DOI 10.12887/36-2023-3-143-06

The aim of the article is to point out a number of reasons why we have the moral right to “tame” artificial intelligence, i.e., to make artificial intelligence subordinate to humans. I attempt to find a justification for this thesis in two main areas. The first area focuses on the negative consequences of applying artificial intelligence in daily life. The second area is related to an attempt to answer the question whether artificial intelligence is a moral agent. The conducted analyses lead to two conclusions. Firstly, artificial intelligence can and does generate negative individual or social consequences. Secondly, artificial intelligence is not a moral agent because it lacks relevant, structural moral properties which human beings possess.

Keywords: artificial intelligence, superintelligence, moral agent, negative consequences

Contact: Department of Ethics, Institute of Philosophy, Faculty of Philosophy,
John Paul II Catholic University of Lublin, Al. Raławickie 14, 20-950 Lublin,
Poland

E-mail: marcin.ferdynus@kul.pl

<https://pracownik.kul.pl/marcin.ferdynus>

ORCID: 0000-0003-0176-1023

Paweł POLAK
Roman KRZANOWSKI

HOW TO TAME ARTIFICIAL INTELLIGENCE? A Symbiotic AI Model for Beneficial AI

In this paper, we posited that in light of the possible risks presented by the uncontrolled development of AI, the original goal of AI that was proposed at a workshop at Dartmouth University in 1956 and then reaffirmed in many subsequent publications should be revised, and a new goal based on the concept of Beneficial AI should be adopted. Furthermore, we proposed that the framework in which we could conceptualize Beneficial AI and its development can be broadly based on the domestication or taming of animals, something that humans have been doing for thousands of years.

The development of artificial intelligence (AI) has been, and continues to be, mostly dominated by engineers, computer scientists, and technological visionaries.¹ The notion of AI and the foundational goals that have chartered its course for decades were conceived at a workshop at Dartmouth University in 1956, with this being attended by John McCarthy, Marvin L. Minsky, Claude E. Shannon, and John Holland.² McCarthy and his colleagues proposed establishing a new discipline called artificial intelligence with the objective of “creating computer systems that could learn, react, and make decisions in a complex changing environment.”³ These original goals were restated and reconfirmed half a century later in 2006 at an event called *The Dartmouth College Artificial Intelligence Conference: The Next 50 Years*.⁴ Current perspectives for the objectives of AI still echo these initial dreams (e.g., “AI is computer science, which aims to develop intelligent machines that can mimic human behavior.”⁵). The

¹ See, e.g., Eugénio Oliveira, “Beneficial AI: The Next Battlefield,” *Journal of Innovation Management* 5, no. 4 (2018): 6–17; Stuart J. Russell, *Human Compatible: Artificial Intelligence and the Problem of Control* (New York: Penguin Books, 2020).

² See John McCarthy, Marvin L. Minsky, Nathaniel Rochester, and Claude E. Shannon, “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence,” August 31, 1955, <http://raysolomonoff.com/dartmouth/boxa/dart564props.pdf>.

³ See Russell, *Human Compatible*.

⁴ See James Moor, “The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years,” *AI Magazine* 27, no. 4 (2006): 87–9.

⁵ See “Future of Artificial Intelligence—Javatpoint,” JavaPoint, 2021, <https://www.javatpoint.com/future-of-artificial-intelligence>.

predictions, however, have become more specific in promising great advances in many areas of life, although not without danger.⁶

Some AI experts go even further in their claims: “The research suggests that our shared AI future is, in part, about our becoming cyborgs. It’s a slow, symbiotic coevolution of our own choosing, and it’s already begun.”⁷ These and other similar predictions have been made by programmers, computer scientists, engineers, businessmen, and journalists, while philosophers, humanists, and other non-technical folk have been conspicuously absent from these discussions. Only sci-fi writers occupying the fringes of science have voiced concerns about the benefits of AI that were spelled out (implicitly) in McCarthy’s manifesto, at least if AI is left untamed in the hands of technocrats.⁸

Over the past few years, however, more and more people beyond the core AI research teams have begun to perceive potential problems with AI,⁹ forcing the AI community to engage in an open discussion about the very nature of this technology. The risks of AI are usually reduced to several clear, well-defined distinct issues, such as job security, safety, privacy, democracy, health, poverty, and so. Yet the real problem with AI is more fundamental, more philosophical rather than related to a specific application or implementation. Indeed, the problem concerns the foundational assumptions behind AI technology itself: Why, and for what, are we developing AI?

Some sixteen years after that workshop in Dartmouth, the goal of AI remains a nuanced version of McCarthy’s 1956 vision of “creating computer systems that could learn, react, and make decisions in a complex changing environment,” although it is now couched in the jargon of current technology.¹⁰

⁶ See, e.g., Katharine Gammon, “5 Ways Artificial Intelligence Will Change the World by 2050,” USC News, December 04, 2017, <https://news.usc.edu/trojan-family/five-ways-ai-will-change-the-world-by-2050/>; Janna Anderson and Lee Rainie, “Artificial Intelligence and the Future of Humans,” Pew Research Center: Internet, Science & Tech (blog), December 10, 2018, <https://www.pewresearch.org/internet/2018/12/10/artificial-intelligence-and-the-future-of-humans/>; Ashley Stahl, “How AI Will Impact The Future Of Work And Life,” Forbes, March 10, 2021, <https://www.forbes.com/sites/ashleystahl/2021/03/10/how-ai-will-impact-the-future-of-work-and-life/>.

⁷ Mike Bectel, “The Future of AI: Seeing the Forest for the Trees, and the Forests Beyond,” Deloitte AI Institute, 2021, <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/process-and-operations/us-ai-institute-future-of-ai.pdf>.

⁸ See Roman Krzakoński and Paweł Polak, “The Future of AI: Stanisław Lem’s Philosophical Visions for AI and Cyber-Societies in Cyberiad,” *Pro-Fil* 22, no. 3 (2021): 39–53.

⁹ “Stephen Hawking, Elon Musk, Steve Wozniak, Bill Gates, and many other big names in science and technology have recently expressed concern in the media and via open letters about the risks posed by AI, joined by many leading AI researchers.” Max Tegmark, “Benefits & Risks of Artificial Intelligence,” Future of Life Institute, 2016, <https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/>.

¹⁰ The terms currently encapsulating AI objectives are Artificial General Intelligence or Human Intelligence (see, e.g., Melanie Mitchell, *Artificial Intelligence: A Guide for Thinking Humans*

Thus, we make the claim that the problem with AI lies in its original manifesto, the fundamental 1956 vision for why we pursue AI; the vision that in fact did not evolve. We also posit that in order to avoid the current and future problems that AI technology is likely to bring, in addition to its numerous benefits, our fundamental vision of AI should shift from those original 1950s objectives to something new.¹¹ The question, though, is to what should it shift?

To avoid or at least limit the risks of AI, we should drop the ambition of building super-intelligent, transhuman systems that mimic or exceed human capacities.¹² In trying to blindly replicate and improve upon human intelligence, according to the 1956 vision, we should heed one of Kant's warnings: "Out of the crooked timber of humanity, no straight thing was ever made."¹³ We should instead focus on building AI systems that will be beneficial to humanity in the most general sense. In other words, AI systems should benefit us not just in a particular field but rather in principle, in essence. Even if it sounds like a cliché, AI systems should be good to us as a technical requirement.

The objective of AI research should be to create AI that benefits us in all walks of life rather than creating a perfect reckoning system, which was McCarthy's original idea for AI. We can label such AI systems as "Beneficial AI," and this is what this paper is all about. The way to do this, we propose, is based on the analogy of taming animals, so we will "tame" AI. The metaphor of taming (or domestication) will be taken in its basic sense, where it influences the relationship between humans and other biological organisms. We therefore work under the assumption that the process of taming an animal can in principle be helpful in bringing about the desired model of Beneficial AI, especially as we are apparently very good at taming wild things.¹⁴

(London: Penguin, 2019); Michael Timothy Bennett, "Computable Artificial General Intelligence," arXiv, May 30, 2022, <http://arxiv.org/abs/2205.10513>; Gary Marcus, "Artificial General Intelligence Is Not as Imminent as You Might Think," *Scientific American*, June 6, 2022, <https://www.scientificamerican.com/article/artificial-general-intelligence-is-not-as-imminent-as-you-might-think/>; Ragnar Felland, "Why General Artificial Intelligence Will Not Be Realized," *Humanities and Social Sciences Communications* 7, no. 1 (2020): 10; Samu Kumpulainen and Vagan Terzian, "Artificial General Intelligence vs. Industry 4.0: Do They Need Each Other?," *Procedia Computer Science* 200 (2022): 140–50).

¹¹ See also Oliveira, "Beneficial AI"; Russell, *Human Compatible*.

¹² For the incorrect focus of our AI-development strategies, see also Russell, *Human Compatible*.

¹³ John Banville, "Foreword," in Isaiah Berlin, *The Crooked Timber of Humanity: Chapters in the History of Ideas*, ed. Henry Hardy (Princeton and Oxford: Princeton University Press, 2013), XI–XVIII, Jstor, <http://www.jstor.org/stable/j.ctt2tt8nd.3>. Kantian concerns about replicating and improving upon the human mind were somewhat reflected by Russell when he stated that "the very definition of success in AI is wrong," meaning that we are developing AI under flawed objectives (see Russell, *Human Compatible*, 13).

¹⁴ See, e.g., Jared M. Diamond, *Guns, Germs and Steel: The Fates of Human Societies* (London: Cape, 1997); Jessica Leary, "Our Furry Friends: The History of Animal Domestication,"

The concept of Beneficial AI is not new, because it has been elaborated upon in several publications.¹⁵ However, the idea of “taming AI” as a means to progress toward Beneficial AI has not been discussed thus far.

WHY DO WE NEED TO REVISE THE GOALS OF AI?

AI has provided many very useful solutions, so it is therefore the technology that is currently being pinned as the greatest hope for improving human life. At the same time, however, the proliferation of AI systems is causing increasing risks.¹⁶ The main philosophical determinants of these threats can be boiled down to four fundamental philosophical problems.

First, people are increasingly losing their decision-making power (i.e., a loss of primary decision agency). Second, they are also gradually losing control over their cognitive processes (i.e., a loss of primary epistemic agency). Third, they are gradually losing control over technical systems and the development of new technology (i.e., a loss of control). Fourth, AI systems have an overwhelming advantage when it comes to processing large amounts of data, allowing them to solve problems more efficiently and gain a significant advantage (i.e., the supremacy of computing power).

The lack of clarity surrounding the risks posed by AI has partly resulted from the peculiar ideology driving the development of AI. The project to create a “synthetic human,” which is an explicit or implicit premise of the AI discussion, appeals to myths and metaphysical longings, and it may even be a challenge to the Creator. However, it should be noted that what is actually being undertaken is the construction of an ideological envelope for the AI program, because it effectively obscures the actual goals and problems by diverting people’s attention to intriguing but irrelevant aspects. After all, the very use of the term “artificial intelligence” is very ideologically loaded, and this is particularly evident when comparing it to the program’s original label, namely cybernetics.

Journal of Young Investigators, February 17, 2012, <https://www.jyi.org/2012-february/2017/9/17/our-furry-friends-the-history-of-animal-domestication>.

¹⁵ See, e.g., “Beneficial AI 2017,” Future of Life Institute, January 12, 2017, 2022, <https://futureoflife.org/bai-2017/>; Oliveira, “Beneficial AI”; Russell, *Human Compatible*; Pedro Fernandes, Francisco C. Santos, and Manuel Lopes, “Norms for Beneficial AI: A Computational Analysis of the Societal Value Alignment Problem,” *AI Communications* 33, nos. 3–6 (2020): 155–71; “Is ‘Provably Beneficial’ AI Possible?” ITU Hub, September 29, 2020, <https://www.itu.int/hub/2020/09/is-provably-beneficial-ai-possible/>.

¹⁶ See, e.g., Tegmark, “Benefits & Risks of Artificial Intelligence.”

It should also be noted here that according to the rationality of technical activities, AI is designed with the aim of achieving specific, often hidden, business goals. This makes it possible to develop the program, although such a choice leads to a reduced axiology for the activities undertaken. More and more voices claim that it is possible to account for other hierarchies of values and goals in the design of technical systems and thus humanize technology, but this has not led to a wider discussion, probably because of the limited awareness of the risks. Instead, a significant problem is a lack of ideas for an AI model that will benefit societies in the long term rather than just profit a narrow group in the short term. In other words, the current business model for AI development has no competitors, and this is a serious problem that needs to be confronted when addressing the threats of AI.

TOWARD BENEFICIAL AI PRELIMINARIES

To organize the philosophical considerations surrounding AI, we should first pose some basic questions. In particular, we should question what we expect from AI:

- (1) Do we need superhumans?
- (2) Do we desire perfect slaves?
- (3) Do we want synthetic humans?
- (4) Is a symbiotic coexistence appealing?

The answers to such questions reveal important differences between philosophers, tech visionaries, and AI researchers and engineers, so we need to clarify some fundamental questions. We should therefore pose some philosophical questions rather than the previous set of questions:

- (1) What kind of AI do we need as humanity?
- (2) What kind of relationships do we need?
- (3) What values should be preferred?
- (4) Is the anthropocentric viewpoint on AI justified?

The first question sets the perspective for the whole deliberation, while the subsequent questions are refinements of this perspective, which we will call *Beneficial AI*. As we understand it, this concept represents the AI that we need as humans for beneficial development in the long term.

BENEFICIAL AI

What is Beneficial AI? Stuart J. Russell's definition states that a beneficent machine, one driven by Beneficial AI, realizes our objectives rather than its

own.¹⁷ Of course, it would be simpler if we knew what we really wanted,¹⁸ but this is not the case. Thus, Russell proposes some tentative guidelines under which beneficially inclined AI systems should operate. He qualifies his proposal by admitting that these are just guidelines rather than rules of any sort, because he fears that these may be taken like Isaac Asimov's notorious laws of robotics, which were originally proposed in Asimov's work *I, Robot*¹⁹ and amended several times. Such an approach risks pushing the whole idea of Beneficial AI down the rabbit's hole.²⁰

Russell's rules for Beneficial AI, which are not indented as laws,²¹ state, firstly, that the machine objective is to maximize the realization of human preferences. Secondly, they assert that the machine does not know initially what these preferences should be. Thirdly, they posit that the machine learns these preferences from human behavior. Russell is fully aware that we do not actually know how to do this, technically, conceptually, or otherwise, but he is sure that if we want to avoid the potential calamities of unbridled AI development, we must pursue this endeavor.

The concept of Beneficial AI has also been elaborated in the Asilomar AI Principles.²² This list of recommendations from the Beneficial AI Conference is a lengthy one,²³ but a few of the more important ones include:

(1) Ethics: AI systems should be designed and operated such that they are compatible with the ideals of human dignity, rights, freedoms, and cultural diversity.

(2) Value alignment: Highly autonomous AI systems should be designed such that their goals and behaviors are guaranteed to align with human values throughout their operation.

(3) Shared benefits: AI technologies should benefit and empower as many people as possible.

¹⁷ See Russell, *Human Compatible*.

¹⁸ See *ibidem*.

¹⁹ See Isaac Asimov, *I, Robot* (Garden City, New York: Doubleday & Company, Inc., 1950).

²⁰ This expression is used especially in the phrase "going down the rabbit hole" or "falling down the rabbit hole." It is a metaphor for something that transports someone into a wonderful (or troublingly) surreal state or situation (see "Rabbit Hole," Dictionary.com, <https://www.dictionary.com/e/slang/rabbit-hole/>). The expression dates back to the famous 1865 classic *Alice's Adventures in Wonderland* by Lewis Carroll (see Lewis Carroll, *Alice's Adventures in Wonderland* Cambridge: Cambridge University Press, 1865), who was less famously a mathematician.

²¹ See Russell, *Human Compatible*, 172.

²² "Asilomar AI Principles," Future of Life Institute, August 11, 2017, <https://futureoflife.org/2017/08/11/ai-principles/>.

²³ See "Beneficial AI 2017."

As we said, the list is long, with it comprising twenty three areas grouped into research issues, ethics and values, and long-term issues.²⁴ These ideas are certainly on the mark, and one could say they benefit the discussion about AI. Anyway, we more or less know what Beneficial AI should be, but the problem is that we are not sure how to realize it. This is where the concept of domestication or taming comes in.

A SYMBIOTIC PERSPECTIVE ON BENEFICIAL AI

An important task for contemporary philosophers should be to search for ways to understand which AI models will meet the generally defined requirements of Beneficial AI. In principle, the task at hand seems practically impossible, since it would require first solving problems that have plagued humanity for centuries. The absence of any reasonable hope for generally solving this problem does not automatically lead to skepticism, though, because this incredibly complicated issue can be simplified in a non-trivial way by imitating a successful strategy from the history of *Homo sapiens*' development. This is admittedly an inductive inference, but the mechanisms of biological cooperation and domestication are still widely used and play an important role in people's lives. So, can such a strategy be used for AI? In other words, is it valid to analogize embodied, biosemiotic AI systems to biological species? The arguments raised below indicate that the proposed approach could be justified, but every approach to AI should be evaluated separately. Whether it is adequate, and to what extent, can only be established through experience. Nevertheless, the problem to be solved is so weighty, and the prospect so promising, that it is worth taking a risk and testing out this theory.

Now, let's try to look at autonomous AI systems as a specific species that coexists in the human environment. The biological perspective then gives, through analogy, a wide range of concepts for describing non-competitive (symbiotic) relationships. It is noteworthy that from such a perspective, the beneficence (i.e., usefulness) of an AI system can be understood as being analogous to that of animals, namely not as the utility of a tool but rather as a beneficial coexistence. Note also that such a biologically inspired perspective offers more possibilities than the model of an artificial slave, which is assumed in Asimov's famous laws of robotics.²⁵

²⁴ See "Asilomar AI Principles."

²⁵ For a critique of Asimov laws' application to robotics, see, e.g., Susan Leigh Anderson, "The Unacceptability of Asimov's Three Laws of Robotics as a Basis for Machine Ethics," in *Machine Ethics*, ed. Michael Anderson and Susan Leigh Anderson (Cambridge: Cambridge University

Symbiotic AI therefore appears to be a very promising model for creating useful AI. Symbiotic relationships are well-tested as a strategy for enhancing the developmental capabilities of organisms, and this approach draws attention to the fact that AI systems are, whether we want them to be or not, part of the operating environment of modern humans. Biological metaphors then make it possible to effectively shape inter-species relations, including those with artificial species, through taming/domestication by introducing an important conceptual and theoretical framework for Beneficial AI concepts.

TAMING AI

AI SYSTEMS: ARTIFACTS OR SPECIES?

Let us start our investigation about how to interpret the AI phenomenon with a brief review of the arguments and counterarguments. The arguments in favor of regarding AI systems as mere tools first assert that they are sophisticated, extremely complex technical artifacts. They also point out that AI systems fulfill roles as tools, as they were intended to. It is also clear that AI systems are a result of typical technical processes, such as design, manufacture, testing, and use. These arguments are generally valid, although they are certainly one-sided. It should also be noted that they are entirely accurate for some AI techniques. The conclusion to be drawn is that the concept of AI is overly broad, and to avoid confusion in further considerations, only a select class of technical solutions that are labelled as AI should be considered. In this paper, we choose to focus on autonomous robotic AI systems (i.e., embodied AI) that have been designed from a biosemiotic perspective. For such systems, we will therefore address the arguments for regarding them as an artificial species.

First, their autonomy excludes them from being treated as typical tools, because tools lack autonomy and independent initiative, so the actions of a human determine whether a tool functions properly. Secondly, the adaptability of these AI models, along with their autonomy, makes the relationship with such artifacts different from that with typical tools, which have a strictly defined purpose, a scope for correct behavior, and no independent decision-making abilities. Thirdly, machines can enter into more complex interpersonal relationships. Even simple and unsophisticated systems like ELIZA have strongly engaged people in peculiar social relationships. Fourth, in such systems, it is possible for them to take over some organizational and social functions (e.g.,

Press, 2011), 285–96; Peter W. Singer, “Isaac Asimov’s Laws of Robotics Are Wrong,” *Brookings*(blog), 2010, <https://www.brookings.edu/opinions/isaac-asimovs-laws-of-robotics-are-wrong/>.

care for the elderly). Finally, we are observing evident bioinspiration in the latest paradigms for AI technology²⁶, such as embodied AI and the biosemiotic paradigm in AI.

It is worth paying some attention to the biosemiotic perspective adopted here.²⁷ This is based on the acceptance of diverse, species-determined forms of intelligence, and it also assumes the possibility of reconstructing biosemiotic relations in artificial systems. Such an AI system will be an embodied autonomous system with its own internally generated field of meaning. The failures of the AGI (artificial general intelligence) concept flow from a misunderstanding of the distinctiveness of the intelligence of an artificial “machine species.”

Incorrect assumptions about AI²⁸ exclude the possibility of understanding the goals of the AI program. Thus, if the goal of AI is not to create an “artificial human,” what are the actual goals?

TAMING AND DOMESTICATION OF ANIMALS

Taming and domestication of animals have been important cultural activities throughout history and facilitated the fascinating development of *Homo sapiens*. Such processes enabled the development of human societies by providing stable sources of food, materials, and energy. Dogs were the first species to be domesticated about fifteen thousand years ago. Another momentous event in human history was the domestication of sheep and goats between 9,000 and 7,000 BC.

Coexistence and cooperation between different, often very different, species are made possible through fundamental biological relationships that can be labelled under the umbrella term *symbiosis*. Symbiosis is defined as a non-antagonistic, long-term relationship, and two basic forms manifest:²⁹

- (1) *mutualism*, where both species gain benefits;
- (2) *commensalism*, where an asymmetric benefit occurs (i.e., where one species benefits while the other species is neither benefitted nor

²⁶ More about AI paradigms, see: Roman Krzanowski and Paweł Polak, “Ontology and AI Paradigms,” *Proceedings* 81, no. 1 (2022): 119. <https://www.mdpi.com/2504-3900/81/1/119>.

²⁷ See Anna Sarosiek, “Biologiczne i semiotyczne nurty w dziedzinie sztucznej inteligencji,” *Transformacje* 1–2, nos. 88–89 (2016): 294–306; Anna Sarosiek, “The Role of Biosemiosis and Semiotic Scaffolding in the Processes of Developing Intelligent Behaviour,” *Philosophical Problems in Science (Zagadnienia Filozoficzne w Nauce)*, no. 70 (2021): 9–44.

²⁸ See, e.g., Brian Cantwell Smith, *The Promise of Artificial Intelligence: Reckoning and Judgment* (Cambridge, MA: The MIT Press, 2019); Michael Wooldridge, *The Road to Conscious Machines: The Story of AI* (London: Penguin, 2021).

²⁹ See Pierre Joseph van Beneden, *Animal Parasites and Messmates* (London: Kegan Paul, Trench, 1883).

harmed—a typical example is how dogs follow a commensal pathway into domestication).

The possibility of analogously applying biological concepts to artificial systems depends upon whether AI can be treated as a synthetic species. This will be the subject of the next part of the article.

TAMING AND DOMESTICATING AI

The prospect of taming synthetic species and integrating them into symbiotic relationships brings many new challenges. First, it is relativized to just a certain group of AI systems, although it should also be noted that there is still no consensual understanding of the concept of AI autonomy. Second, the ability to design the traits of artificial systems means the relationship can be developed more quickly than with biological organisms, where generations of artificial selection may be required. This is where one of the critical problems arises: The wide scope of feature design and the ability to design quickly makes the time required for reception (i.e., adaptation, achieving cooperation) a critical factor. In this respect, taming AI and creating a symbiotic relationship with it will be incomparably more difficult than it is for biological species.

It is also worth raising questions about the specific subjective conditions for taming such an “artificial species.” Recall that we refer here to an embodied AI equipped with its own internally induced semantics.

The most important propositions for such conditions can be summarized as follows:

- (1) social acceptance of AI as a separate species (IT worldview)³⁰;
- (2) a lack of psychological barriers (e.g., uncanny valley);
- (3) no self-reduction of humans, where relationships are shaped like relationships with a tool rather than a living organism (Bolter’s thesis)³¹;
- (4) a need for critical thinking to accelerate cultural evolution (a critical time factor!);
- (5) openness to change in the image of the world (i.e., the image of humanity, machines, society).

³⁰ See Witold Marciiszewski and Paweł Staciewicz, *Umysł—Komputer—Świat. O zagadce umysłu z informatycznego punktu widzenia* (Warszawa: Akademicka Oficyna Wydawnicza EXIT, 2011), <http://libra.ibuk.pl/book/101353>.

³¹ See Jay David Bolter, *Turing’s Man: Western Culture in the Computer Age* (Chapel Hill: University of North Carolina Press, 1984).

*

In this paper, we posited that in light of the possible risks presented by the uncontrolled development of AI, the original goal of AI that was proposed at a workshop at Dartmouth University in 1956 and then reaffirmed in many subsequent publications should be revised, and a new goal based on the concept of Beneficial AI should be adopted. Furthermore, we proposed that the framework in which we could conceptualize Beneficial AI and its development can be broadly based on the domestication or taming of animals, something that humans have been doing for thousands of years.

We have to recognize that taming of animals is a process different in kind, not in a degree, from taming of AI systems. In fact, apart from broad similarities, the details how to tame AI systems are not immediately obvious. The details are not obvious as in taming of AI we do not have the experience of thousands of years of taming animals and other human beings to inform us. There is a lot, if not everything, to learn how to do it right.

The issue of taming AI also opens up interesting fields of ethical research, for example: can the complex human–AI symbiotic relationships even be meaningfully described by utilitarian ethics without falling into unacceptable simplifications? What kind of ethics would be appropriate for such an analysis?³² Clarifying the concepts involved in ethical issues of taming of AI is certainly necessary to responsibly address such questions. This should become the subject of future studies.

AI systems are an important part of the widely understood human environment, so it is necessary to rethink the goals of AI development. Indeed, this is one of the most important tasks facing philosophy today (i.e., philosophy in technology). As we indicated, symbiotic relationships could be used to shape useful AI. The analysis shows that the critical factor in the face of differentiation in AI is the reception time factor, because differentiated AI systems makes it impossible to undertake a general analysis to find general solutions. The requirements for AI systems in symbiotic relationships must therefore be formulated for a specific context.

On the other hand, it is possible to formulate some general subject requirements for humans in these symbiotic relationships. It is imperative to develop the emerging concept of Beneficial AI, despite its numerous difficulties, due to its critical importance for humans.

Such an approach admittedly does not solve the philosophical problems associated with the development of AI, but it quite effectively diverts attention

³² The authors wanted to thank the anonymous reviewer for bringing these issues to their attention.

from them. On the other hand, the numerous dangers and threats associated with the rapid development of AI technology, both real and imaginary, bring some important philosophical themes back into the discussion.

The progressive technologization of individual and social life makes humanizing the technological sphere a particularly important issue. The task of philosophy is therefore to set goals for our rapidly developing technology, so it will serve the long-term interests of modern human life, in contrast to the hitherto increasingly pronounced dehumanization process. If this task is not undertaken, the future will be shaped along the lines of a technocratic project, or it will be susceptible to strong irrational elements hidden within uncritical visions of technological development and fueled by wishful thinking.

BIBLIOGRAPHY / BIBLIOGRAFIA

- Anderson, Janna, and Lee Rainie. "Artificial Intelligence and the Future of Humans." Pew Research Center: Internet, Science & Tech (blog), December 10, 2018. <https://www.pewresearch.org/internet/2018/12/10/artificial-intelligence-and-the-future-of-humans/>.
- Anderson, Susan Leigh. "The Unacceptability of Asimov's Three Laws of Robotics as a Basis for Machine Ethics." In *Machine Ethics*. Edited by Michael Anderson and Susan Leigh Anderson. Cambridge: Cambridge University Press, 2011.
- "Asilomar AI Principles." Future of Life Institute, August 11, 2017. <https://futureoflife.org/2017/08/11/ai-principles/>.
- Asimov, Isaac. *I, Robot*. Garden City, New York: Doubleday & Company, Inc., 1950.
- Banville, John. "Foreword." In Isaiah Berlin, *The Crooked Timber of Humanity: Chapters in the History of Ideas*. Edited by Henry Hardy. Princeton and Oxford: Princeton University Press, 2013. <http://www.jstor.org/stable/j.ctt2tt8nd.3>.
- Bechtel, Mike. "The Future of AI: Seeing the Forest for the Trees, and the Forests Beyond." Deloitte AI Institute, 2021. <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/process-and-operations/us-ai-institute-future-of-ai.pdf>.
- van Beneden, Pierre Joseph. *Animal Parasites and Messmates*. 3rd ed. London: Kegan Paul, Trench, 1883.
- "Beneficial AI 2017." Future of Life Institute, January 12, 2017. <https://futureoflife.org/bai-2017/>.
- Bennett, Michael Timothy. "Computable Artificial General Intelligence." arXiv, May 30, 2022. <http://arxiv.org/abs/2205.10513>.
- Bolter, Jay David. *Turing's Man: Western Culture in the Computer Age*. Chapel Hill: University of North Carolina Press, 1984.
- Diamond, Jared M. *Guns, Germs and Steel: The Fates of Human Societies*. London: Cape, 1997.

- Dictionary.com (s.v. "Rabbit Hole"). <https://www.dictionary.com/e/slang/rabbit-hole/>.
- Fernandes, Pedro, Francisco C. Santos, and Manuel Lopes. "Norms for Beneficial AI: A Computational Analysis of the Societal Value Alignment Problem." *AI Communications* 33, nos. 3–6 (2020): 155–71.
- Fjelland, Ragnar. "Why General Artificial Intelligence Will Not Be Realized." *Humanities and Social Sciences Communications* 7, no. 1 (2020): 10.
- Gammon, Katharine. "5 Ways Artificial Intelligence Will Change the World by 2050." USC News, December 04, 2017. <https://news.usc.edu/trojan-family/five-ways-ai-will-change-the-world-by-2050/>.
- "Future of Artificial Intelligence—Javatpoint." JavaPoint, 2021. <https://www.javatpoint.com/future-of-artificial-intelligence>.
- "Is 'Provably Beneficial' AI Possible?" ITU Hub, September 29, 2020. <https://www.itu.int/hub/2020/09/is-provably-beneficial-ai-possible/>.
- Krzanowski, Roman, and Paweł Polak. "The Future of AI: Stanisław Lem's Philosophical Visions for AI and Cyber-Societies in *Cyberiad*." *Pro-Fil* 22, no. 3 (2021): 39–53.
- . "Ontology and AI Paradigms." *Proceedings* 81, no. 1 (2022): 119. <https://doi.org/10.3390/proceedings2022081119>.
- Kumpulainen, Samu, and Vagan Terziyan. "Artificial General Intelligence vs. Industry 4.0: Do They Need Each Other?" *Procedia Computer Science* 200 (2022): 140–50.
- Lear, Jessica. "Our Furry Friends: The History of Animal Domestication." *Journal of Young Investigators*, February 17, 2012. <https://www.jyi.org/2012-february/2017/9/17/our-furry-friends-the-history-of-animal-domestication>.
- Marciszewski, Witold, and Paweł Stacewicz. *Umysł—Komputer—Świat: O zagadce umysłu z informatycznego punktu widzenia*. Warszawa: Akademicka Oficyna Wydawnicza EXIT, 2011. <http://libra.ibuk.pl/book/101353>.
- Marcus, Gary. "Artificial General Intelligence Is Not as Imminent as You Might Think." *Scientific American*, June 6, 2022. <https://www.scientificamerican.com/article/artificial-general-intelligence-is-not-as-imminent-as-you-might-think1/>.
- McCarthy, John, Marvin L. Minsky, Nathaniel Rochester, and Claude E. Shannon. "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence," August 31, 1955. <http://raysolomonoff.com/dartmouth/boxa/dart564-props.pdf>.
- Mitchell, Melanie. *Artificial Intelligence: A Guide for Thinking Humans*. London: Penguin, 2019.
- Moor, James. "The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years." *AI Magazine* 27, no. 4 (2006): 87–91.
- Oliveira, Eugénio. "Beneficial AI: The Next Battlefield." *Journal of Innovation Management* 5, no. 4 (2018): 6–17.
- "Rabbit Hole." In Dictionary.com. <https://www.dictionary.com/e/slang/rabbit-hole/>.

- Russell, Stuart J. *Human Compatible: Artificial Intelligence and the Problem of Control*. New York: Penguin Books, 2020.
- Sarosiek, Anna. "Biologiczne i semiotyczne nurty w dziedzinie sztucznej inteligencji." *Transformacje* 1–2, nos. 88–89 (2016): 294–306.
- . "The Role of Biosemiosis and Semiotic Scaffolding in the Processes of Developing Intelligent Behaviour." *Philosophical Problems in Science (Zagadnienia Filozoficzne w Nauce)*, no. 70 (2021): 9–44.
- Singer, Peter W. "Isaac Asimov's Laws of Robotics Are Wrong." Brookings (blog), 2010. <https://www.brookings.edu/opinions/isaac-asimovs-laws-of-robotics-are-wrong/>.
- Smith, Brian Cantwell. *The Promise of Artificial Intelligence: Reckoning and Judgment*. Cambridge, MA: The MIT Press, 2019.
- Stahl, Ashley. "How AI Will Impact The Future Of Work And Life." Forbes, March 10, 2021. <https://www.forbes.com/sites/ashleystahl/2021/03/10/how-ai-will-impact-the-future-of-work-and-life/>.
- Tegmark, Max. "Benefits & Risks of Artificial Intelligence." Future of Life Institute, 2016. <https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/>.
- Wooldridge, Michael. *The Road to Conscious Machines: The Story of AI*. London, UK: Penguin, 2021.

ABSTRACT / ABSTRAKT

Paweł POLAK, Roman KRZANOWSKI, How to Tame Artificial Intelligence? A Symbiotic Model for Beneficial AI

DOI 10.12887/36-2023-3-143-07

The paper presents a vision of tamed AI, an AI technology that would maximize its benefits to society and individuals. The need for such a technology has been discussed at length under the umbrella term of Beneficial AI, yet there have been no proposals for how to actually implement such a system. In this paper, we discuss the concept of domesticating (or taming) AI, how this would apply to AI technology, and how it could lead to a symbiotic human—machine system. We refer to this proposed AI concept as a symbiotic AI model for Beneficial AI.

Keywords: tamed AI, Beneficial AI, symbiotic systems, symbiotic AI, future of AI, AI objectives, mitigating AI risks

Contact: Department of History and Philosophy of Science, Faculty of Philosophy, Pontifical University of John Paul II, ul. Kanonicza 9, 31-002 Cracow, Poland

E-mail: (Paweł Polak) pawel.polak@upjp2.edu.pl; (Roman Krzanowski) wf@upjp2.edu.pl
(Paweł Polak) <http://polak.wikidot.com/>; (Roman Krzanowski) https://www.researchgate.net/profile/Roman_Krzanowski
ORCID: (Paweł Polak) 0000-0003-1078-469X; (Roman Krzanowski) 0000-0002-8753-0957

Paweł POLAK, Roman KRZANOWSKI – Jak oswoić sztuczną inteligencję? Model symbiotycznej SI jako realizacja dobroczynnej SI

DOI 10.12887/36-2023-3-143-07

W niniejszym artykule zaprezentowano wizję oswojenia sztucznej inteligencji (SI jako techniki, która maksymalizuje korzyści płynące dla używających jej zarówno społeczności, jak i jednostek). Dotychczas o potrzebie takiej techniki dyskutowano, używając pojęcia dobroczynnej SI (ang. beneficial AI), choć wciąż nie pojawiły się konkretne propozycje, w jaki sposób zaimplementować taki system. W opracowaniu zaproponowano wykorzystanie analogii wybranych klas systemów SI do organizmów żywych, co pozwala na rozważenie możliwości udomawiania lub wdrażania takich systemów SI. Strategia domestykacji lub osławiania jest wypróbowaną metodą przynoszącą długofalowe korzyści gatunkowi ludzkiemu, dlatego w artykule podjęto kwestię, w jaki sposób można zastosować tę koncepcję w technice SI oraz jak należy projektować symbiotyczne systemy człowiek–maszyna. W rezultacie sformułowano symbiotyczny model SI dla dobroczynnej SI.

Słowa kluczowe: oswojona SI, dobroczynna SI, systemy symbiotyczne, symbiotyczna SI, przyszłość SI, cele AI, ograniczanie ryzyka związanego z AI

Kontakt: Katedra Historii i Filozofii Nauki, Wydział Filozofii, Uniwersytet Papieski Jana Pawła II, ul. Kanonicza 9, 31-002 Kraków
E-mail: (Paweł Polak) pawel.polak@upjp2.edu.pl; (Roman Krzanowski) wf@upjp2.edu.pl
(Paweł Polak) <http://polak.wikidot.com/>; (Roman Krzanowski) https://www.researchgate.net/profile/Roman_Krzanowski
ORCID: (Paweł Polak) 0000-0003-1078-469X; (Roman Krzanowski) 0000-0002-8753-0957

SZTUCZNA INTELIGENCJA W SPOŁECZEŃSTWIE LUDZKIM

Mariusz WOJEWODA

ARTIFICIAL INTELLIGENCE AS A SOCIAL UTOPIA

AI we can understand as the logic of machine thinking, which constitutes a foundation for the introduction of optimal operating principles for the smooth functioning of the system. This provides the context for the author's analysis of the logic of power in two aspects: (1) AI as a model for exercising control over members of institutions; (2) the possibility of exercising control over AI by humans. In both cases, we are dealing with a variant of the social utopia, the assumption being that such control brings the desired effects, namely, an increase in the efficiency of the system involving people and machines.

When we talk about post-humanity, we should always be attentive to how we understand humanity itself. Perhaps, the prospect of post-humanity will enable us precisely to gain a new insight into what being-human means.¹

Slavoj Žižek

INTRODUCTION: TECHNO-UTOPIA AS A CURRENT ISSUE?

Artificial intelligence can be analyzed in two aspects: (1) as a technical solution, and (2) as the impact of AI artifacts on man. In the first case, the problem of AI consists in creating machines that to some extent can function independently of human control. In the second variant, AI is about creating a system that will complement or replace human agency, a system that will support those aspects of human thinking which involve computation. The latter variant concerns especially operations which involve the ordering, selection, and interpretation of large amounts of data.² Both these approaches to AI posit it as a tool with which to create a utopian vision of a “better” world.

The desire to create a better world is one of the key aspects of utopian social programs, past and present. We tend to turn our attention to utopias created during the Renaissance—Thomas More's *Utopia*,³ Tommaso Campanella's

¹ Slavoj Žižek, *Hegel in a Wired Brain* (London: Bloomsbury Publishing, 2020), 23.

² See David Stephenson, *Big Data Demystified: How to Use Big Data, Data Science and AI to Make Better Business Decisions and Gain Competitive Advantage* (London: Pearson Education, 2018), 32–34.

³ See Thomas More, *Utopia*. Translation and introduction Clarence H. Miller (New Haven and Yale: Yale University Press, 2001).

Civitas Solis,⁴ and Francis Bacon's *New Atlantis*.⁵ The term "utopia" became common thanks to the Latin title of More's essay. The word "utopia" comes from the Greek word *ou-topos* and means a place that does not exist; alternatively, it derives from the word *eu-topos* and means "a good place." Understood in the latter sense, utopia denotes a state of existence which is regarded as "better" in relation to the previous state and this meaning is applicable in the context of the issue of techno-utopia.⁶ The creators of Renaissance political utopias assumed that the improvement of human life resulted from the change of the political system, but also from the development of scientific knowledge and the use of technical tools. Modern utopias fit into this tendency to admire the possibilities offered by modern technical inventions. By analogy with the utopian symbolism of the "better place," one can speak of a "more perfect," more effective functioning of machines and man. Admiration for technical inventions fits into a broad current of concepts referred to as transhumanism or posthumanism. The most famous representatives of this trend are Ray Kurzweil,⁷ Nick Bostrom,⁸ and Susan Schneider.⁹ Alongside them, as it were, there are also theories put forth by those researchers—Shoshana Zuboff¹⁰ and Slavoj Žižek¹¹, to name two who analyze the impact of machine technological thinking on the functioning of human societies. They stress that technologization occurs to the extent to which humans "imitate" artificial intelligence to optimize human agency.

The technological utopia is a contemporary variant of the social utopia and can be inscribed in the techno-evolutionary perspective.¹² *Techne* refers to a combination of man-made tools and human skills into a system of technology. Here, however, the danger emerges that techno-utopia, aimed at improving the quality of human life, will turn into a dystopia. This concern points to negative consequences of introducing changes related to the use of AI in the hope of

⁴ See Tommaso Campanella, *Civitas Solis*. Translated by Daniel J. Donno (Berkeley and Los Angeles: University of California Press, 1981).

⁵ See Francis Bacon, *New Atlantis* (Cambridge: Cambridge University Press, 2013).

⁶ See Jerzy Szacki, *Spotkania z utopią* (Warszawa: Państwowy Instytut Wydawniczy, 1980), 10.

⁷ See Ray Kurzweil, *The Singularity Is Near: When Humans Transcend Biology* (London: Penguin Publishing Group, 2006).

⁸ See Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford: Oxford University Press, 2014),

⁹ See Susan Schneider, *Artificial You: AI and the Future of Your Mind* (Princeton and Oxford: Princeton University Press, 2019).

¹⁰ See Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (New York: Perseus Books, 2019).

¹¹ See Žižek, *Hegel in a Wired Brain*.

¹² See "Technological Utopianism," in Wikipedia, https://en.wikipedia.org/wiki/Technological_utopianism.

increasing the effectiveness of activities of individuals and groups of people in institutions. One scenario considers a possible loss of control over robots equipped with AI or IT systems which will be given the right to decide about their future. It is also necessary to be aware of threats related to the use of artificial intelligence as an element of control over one group of people by another group of people, in which case we are looking at a dehumanizing use of AI as a means to create a technologically-processed collective human intelligence. The imaginary structure of such an intelligence is based on known and approved values, such as safety, effectiveness, accumulation of knowledge, progress, easy access to information, and an ability to extend and improve the quality of human life. One can put forward the thesis that, by implementing these values in a systemic connection with AI, we will have to accept the loss of personal individuality and the ability to decide about ourselves.

Compared to previous ones, contemporary techno-utopias do not have to resort to the means of external oppression. They replace physical violence with system of control, or rather with technical systems that discreetly “observe” users of digital tools, suggest the most accurate solutions, optimize the execution of operations, and enable “discreet” supervision. The resources available to large corporations, e.g., Microsoft, Google, Facebook, make possible the financing of such schemes, which, over time, may become indispensable in the functioning of all major institutions.

The author of this article adopts a hermeneutical perspective, its aim being a critical reflection on the products of technology and their meaning for man. Philosopher of technology Val Dusek has used the term “system of technology” to describe the relationship between man and machine. Technology understood as a system is “the application of scientific or other knowledge to practical tasks by ordered systems that involve people and organizations, productive skills, living things, and machines.”¹³ A hermeneutical approach to the “system of technology” consists in the analysis of relations between man and the artifacts of technology. This approach is concerned with such aspects of these relations as knowledge, inventions, scientific centers, operators, and conservators of machines, public relations specialists and journalists responsible for disseminating knowledge, buyers of technical equipment, and people managing corporations and small companies.¹⁴ The term “system of technology” allows us to describe the complex relationships that occur between man and his products, and to analyze the changes in the image of the world and human

¹³ Arnold Pacey, *The Culture of Technology* (Oxford: Blackwell Publishing Ltd., 1983), 18, <https://www.fulcrum.org/epubs/3x816q10j?locale=en#page=18>. See Val Dusek, *Philosophy of Technology: An Introduction* (Oxford: Blackwell Publishing Ltd., 2006), 35.

¹⁴ See Stephen J. Klein, “What is Technology?”, *Bulletin of Science, Technology & Society*, no. 5 (1985): 215–218.

behavior that occur in connection with it. Currently, artifacts equipped with AI create a specific lifestyle, which is also becoming a model of thinking. In this way, new ways of self-interpretation and of the self-identification of the human subject are emerging, namely, ones in which the artifacts of technology play an essential role. Contemporary narratives about a meaningful and fulfilled life are being related to technology.

To domesticate techne is to make technical tools human-friendly, to introduce them into the human environment. If we treat AI as a new species, we can look for an analogy between the domestication of plants and the taming of animals and the domestication of machines. We assume that man is the domesticator while the machine is an object subjected to domestication, i.e., introduced into the human environment. The word “to domesticate” is defined as the “process of bringing under human control”¹⁵ objects (plants and animals) which are subjected to this process. Hence the connection between the perception of objects and the competence of man, who uses them in the way he sees fit and exercises control over them. The domestication of plants and the taming of animals have taken tens of thousands of years. The question is how much time we will need to domesticate AI and what skills will this process require.

Techno-utopian projects posit the following: (1) unambiguously defined operations leading to the implementation of the utopian project; (2) a clear vision of a new, better state of existence (compared to the previous one) consequent upon the introduction of technical improvements; (3) consistent pursuit of the devised enterprise, regardless of the difficulties; (4) lack of criticism in the pursuit of goals; (5) postulating changes at the level of the functioning of social and political institutions; (6) a program of creating a “better” world “from scratch.”¹⁶ This general characterization translates into specific programs of change, in accordance with the guiding idea associated with the social philosophy derived from the thought of Karl Marx,¹⁷ according to which techno-evolution is not about understanding reality, but above all about changing it. In this context, technical tools equipped with AI lead to modifications of human behavior. The “domestication” of new technologies leads to inevitable changes in the functioning of man. Not all of these changes have negative consequences. The process of domestication does not concern solely the entity subjected to it; it also affects the user, the one who domesticates.

¹⁵ See “Domestication,” in Cambridge Dictionary, Cambridge University Press & Assessment, <https://dictionary.cambridge.org/pl/dictionary/english/domestication>.

¹⁶ See Frank E. Manuel and Fritzie P. Manuel, *Utopian Thought in the Western World* (Cambridge: Belknap Press, 1979), 20–23.

¹⁷ See Andrzej Walicki, *Marksizm i skok do królestwa wolności: Dzieje komunistycznej utopii* (Warszawa: Wydawnictwo Naukowe PWN, 1996), 78n.

AI AS A WAY OF EXERCISING POWER

On the threshold of modern philosophy, Francis Bacon linked domination to knowledge. Having knowledge about the world gives man an advantage over it. Exercising power over “wild” nature and animals turned them into objects subordinated to man.¹⁸ It is doubtful whether the relationship of subordination and control will look similar in the relationship between man and AI, and whether it may lead to the loss of man’s control over his own life. Having the advantage of knowledge over man, AI may gain a dominant position, by analogy with the situation of the advantage that man now has over animals and plants. In cultural narratives—in literature and film—we often see images expressing the fear of losing control over AI artifacts.

In the modern approach, the domination resulting from knowledge means controlling the flow of information, selecting it, organizing it, creating a socially approved model of data interpretation. In one of the possible scenarios, AI becomes a tool in an IT system that allows this system to exercise control over an institution’s employees and customer needs in a way similar to the methods of data management.¹⁹

Let us examine the argument put forward by Shoshana Zuboff. In her opinion, the modern form of exercising power in institutions is not external coercion, but the creation of rules of internal coercion by producing a specific behavioral surplus. Used as a tool of control, AI is not directly interested in us, but in our behaviors and choices. It is focused on the analysis of measurable factors, factors that can be parameterized and made available for the operations of evaluative rendering, calculating, modifying, scoring, and monetizing. The aim of these operations is to maximize the work efficiency of teams and the chances of beating competition. A new form of totalitarianism emerges here, which consists in the generation of rules of self-control in the supervised human subjects.²⁰ The resulting loss of freedom becomes a consequence of the realization of the individual’s own desire for a comfortable and safe life. The successive stages of the realization of these desires turn us into entities who live in a state of prosperity, with no need to make an effort to decide about ourselves or to take responsibility for our decisions.

In a world of mechanical control, there is no need for external supervision, because we watch over ourselves. In this way a new manner of exercising

¹⁸ See Francis B a c o n, *Selected Writings of Sir Francis Bacon* (New York: Franklin Library, 1982), 105.

¹⁹ See Joanna K a m i ń s k a, *Nowe wspaniale światy: Współczesne projekty doskonałego społeczeństwa* (Kraków: Nomos, 2012), 36–38.

²⁰ See Shoshana Z u b o f f, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (New York: Perseus Books LLC, 2019), 226n.

power is created, which Zuboff calls instrumentalism and defines as the adjustment of the fulfillment of the desires of individuals to the realization of the goals of an institution.²¹ This type of totalitarianism has a “soft” and friendly form; it does not negate the value of privacy but treats personal information as input for data processing. The human subject is being controlled but does not rebel because they are not aware of the danger. Chance and uncertainty as to the consequences of human choices have been eliminated. This is like a return to “paradise,” except that paradise now has a new technological-digital form. This paradise should be understood in terms of a way of life and a way of thinking, thus as a utopian no-place. The functions of the host are performed by a friendly system architect.²² By offering to satisfy our needs, the system also relieves us of the hassle of making choices.

For Bacon, the acceptance of two interrelated values, freedom and knowledge, was associated with human agency. The latter value that of agency, was based on the possibility of human influence on the outside world and a person’s power to decide freely. The question of free will plays a fundamental role in the concept of man as the moral subject of his own actions. Behavioral theory, emerging from Burrhus F. Skinner’s psychology of cognition, is based on the negation of the assumption of free will, and consistently puts knowledge in opposition to freedom. In this view, the task of knowledge is to free man from the illusion of freedom. By becoming aware of those areas of activity in which our actions are determined by external causes, we become calmer. The personal freedom of the individual is treated as a cognitive error, and the task of knowledge is to make us aware of the sources of this illusion.²³ Biological factors and the social environment determine human behavior. Ignorance as to how this process occurs breeds fantastical ideas about the freedom and individual agency of the individual. Skinner postulated that we should shift our scientific interests from the analysis of free will to the search for the mechanisms which regulate social behavior. The word “mechanism” indicates a connection between biological and technical factors, factors for which the human will is a “tool” for action and achievement of set goals.

In behaviorist terms, the identity of an individual is a product of factors external to it. The behavior of a biological organism is what another organism sees when observing its activity. Objective descriptions of behavior lead to the discovery of mechanical patterns of action and, ultimately, to the identification of causal relationships between the external environment and the behavior of

²¹ See Zuboff, *The Age of Surveillance Capitalism*, 511; Burrhus F. Skinner, *Science and Human Behavior* (New York: The Free Press, 1965), 85.

²² See Michael Fleischer, *Kapitalizm i jego sztuczna inteligencja* (Wrocław and Kraków: Wydawnictwo Libron, 2022), 40–42.

²³ See Burrhus F. Skinner, *About Behaviorism* (New York: Vintage Books, 1976), 14–16.

individuals and social groups. All behaviors can only be explained from the perspective of the “other” an external observer. Hence the postulate, formulated by Skinner and his continuators, to develop a science of behavior. This postulate concerns, among other things, activities such as problem solving and making moral choices. The study of human behavior takes the form of a conscious biological determinism.

Consequently, technological determinism can be seen as a corollary of biological determinism, considering the way the “technical” basis of human behavior is defined. Zuboff sees in Skinner’s theory a psychological basis of the contemporary techno-utopia, its purpose being to define the rules of human action that can be technically modified. Contemporary behavioral engineering technologies include organizational systems and algorithmic procedures created to generate maximum efficiency for machines and humans. What we are looking at here is a transition from a description that indicates that human behavior can be compared to a machine propelled by external factors to the postulate of transforming this behavior according to the model of a smoothly functioning machine. Cultural affirmation of such values as the dignity of the human person, freedom of choice, and the recognition of the psychophysical distinctiveness of individual persons stand in the way of the automation of human agency.²⁴

Implemented in the spirit of Skinner’s behaviorist psychology, techno-utopia turns into a dystopia. It envisions a world of universal equality, dispassionate harmony, security, an abundance of material goods, and a specific brotherhood, one based on a biological-mental community of consumers. In this intelligent system of behaviors, groups of employees function as teams, producing internal mechanisms of self-control and effective operation. Institutions whose functioning is based on the model of AI are expected to guarantee victory over competitors. The social proof of rightnesses actions will reinforce this kind of behavior. This is manifested in the belief that in order to maintain its position on the market, an institution must adjust its mode of operation according to AI posited as a model; it must adopt it as the optional mechanical model of management.

Zuboff identifies the dangers of techno-utopia by relating it directly to the economic model which she calls, after Skinner, “surveillance capitalism.”²⁵ This suggests that, at least theoretically, there is an alternative socio-economic system, a contemporary form of socialism of sorts that could change this state of affairs. The historical experience of real socialism in the twentieth century

²⁴ See Zuboff, *The Age of Surveillance Capitalism*, 232n.

²⁵ See Burrhus F. Skinner, *Walden Two: With a New Preface by Author* (Cambridge: Hackett Publishing Company, INC, 2005), 173. Quoted after Zuboff, *The Age of Surveillance Capitalism*, 504n.

indicates that socialism can also take the form of totalitarianism. In my opinion, contemporary socialism would also use AI tools. Zuboff's use of the term "capitalism" in the phrase "surveillance capitalism" must be treated with caution; at the same time, the author persuasively argues that an analysis of the impact of AI on social life is not about the technology itself, but about the logic of AI. This logic cannot be realized without modern digital technologies, which may not be directly responsible for it. Modern AI-supported techne is a tool that institutions use to optimize their exercise of power and control over the flow of information.²⁶ At present, there is no convincing model of economic life as an alternative to capitalism.

AI-based managing style refers to the governance model of institutions, but is also about power over people's imagination and future-oriented projects. This project is based on the idea of creating a real-digital world (this combination of two ontological dimensions is intentional) in which privacy of each of us will be digitized and planned. Machine (artificial) intelligence will be a tool used to achieve this goal, that of a freedom-deprived world built on the principles of voluntary acceptance.²⁷ This voluntary "transfer" of power to the artificial sovereign will take place peacefully, without the kind of dramatic coups envisioned in SF films. We may not even see this happen. The problem of the tension between freedom and enslavement will become incomprehensible. Bacon's postulate of elevating man to the position of a ruler over the natural world and the world of things will be reversed by accepting the lack of "understanding" and the "inability to control" the artifacts we use.

Žižek arrives at conclusions similar to those drawn by Zuboff. While Žižek's *Hegel in a Wired Brain* is primarily devoted to an analysis of the problem of the human brain linked to a machine, he also deals in this book with issues related to AI. He takes up the interesting issue of collective intelligence. An IT system connected to many devices can accumulate and use the potential of intelligent machines within the network. The question that arises is whether human brains would be able to imitate such a model. Collective human intelligence would be much more efficient than the scattered intelligences of many people, who have to expend a great deal of effort and time to share the knowledge they possess.

Žižek poses an interesting problem: What will happen to the human spirit (culture) if we realize the postulate of collective mechanical intelligence? Here we have in mind the concepts of Singularity. (a term used by Kurzweil) in which the thinking of an individual will become part of the new form of the

²⁶ See Zuboff, *The Age of Surveillance Capitalism*, 18n.

²⁷ See *ibidem*, 66.

idea (borrowed from Georg Wilhelm Friedrich Hegel²⁸) of the objective spirit. It is impossible to imagine the implementation of this postulate without AI. If the “Other” that Skinner wrote about is an intelligent machine, then at some point “everyone” will think like an intelligent machine. Looking at our reflection in the “artificial” Other, we become “copies.” Such “copies,” equipped with up-to-date and complete information, will be more operative and effective than biological people thinking separately. According to Žižek, Hegel’s objective spirit will attain realization as a space for universal media communication.

Technological progress, which is a form of self-improvement (transcendence), derives from the human desire to create something more perfect than man.²⁹ On the one hand, this is related to admiration for the genius of the creator; on the other, it leads to a decrease in the creator’s self-esteem. Creation in the area of *techne* makes us aware of biological limitations. In a dimension other than mortality and suffering, it reveals the contingency of human existence. Not being perfect enough, we are fascinated by what we have created, for it is “better” than us. This fascination is enhanced by the fact that we can use AI-equipped artifacts to improve human capabilities in both the personal and collective dimensions of life.

For now, we do not want to become artificial intelligence; yet we do want to use its capabilities to gain an advantage over other people. A person equipped with modern technical devices dominates those who are deprived of them or cannot avail themselves of the artifacts of digital technology. Consequently, the more we want to gain an advantage over others, the more we become dependent on the tools that enable us to be more competitive. According to Žižek, the competition for better opportunities on the labor market, a higher social position, and economic advantage will cause the technicalization of human thinking to increase progressively. At present, we surround ourselves with *techne* artifacts; however, over time, they will become part of our bodies and brains. The evolutionary change of man into a posthuman entity (Singularity) may occur smoothly and imperceptibly. However, the ontic improvement in the quality of functioning in a competitive society has its price. Individual will is an unnecessary burden for a well-functioning human technical intelligence. Soon, we will be able to use more and more devices, structures, and models of people management which will work independently of the factors of volition and free will.³⁰

²⁸ See Georg Wilhelm Friedrich Hegel, *Fenomenologia ducha* (Warszawa: Fundacja Aletheia, 2002), 209.

²⁹ See Žižek, *Hegel in a Wired Brain*, 24n.

³⁰ See *ibidem*, 28–30.

This process of human change is conspicuous in the context of the creation of algorithms which use AI as catalysts for social and economic development (Business Intelligence).³¹ In the area of information management, AI is seen as a technique based on the assumption of improving the efficiency of collecting and selecting information of high volume, variability, and diversity (Big Data). As the amount of data is beyond the capabilities of the human mind and even the combined capabilities of many human minds, information processing and management require introducing machine “thinking” techniques. In a general sense, what we are talking about here is transforming data into information, and information into knowledge. At the same time, we still assume that this knowledge must be understood and interpreted by man. At the next stage of the development of “machine learning” techniques, knowledge acquired in this manner will be incomprehensible to man, the creator of the machine. At that stage, the “thinking of the machine” will become something non-transparent, even magical for us, thus inspiring in us religious respect and admiration, combined with fear of the power of what we do not understand; in other words, it will be the contemporary version of *mysterium tremendum* and *fascinosum*.

However, the techno-utopian vision of the future depicted by Zuboff and Žižek does not have to be so pessimistic. We need time to become used to the artifacts of technology. The young tame these artifacts relatively faster than their seniors and the elderly. Progress in information technology is dynamic and fast-paced, while taming techne products equipped with AI requires time and the acquisition of new skills by the user. In addition to IT, digital and communication competences, there are also those related to the competent use of AI.

In the conclusion of her book, Zuboff refers to the consciousness of the “inhabitants” of the digital world. She argues that the most important thing for us today is to awaken in ourselves the desire to defend the right to maintain our own separateness and the ability to decide about ourselves. This refers to the sense in which values such as the authenticity of reactions, our own emotions remain important to us. If we do not want our statements to be digitally cataloged, manipulated, and then used for commercial or ideological purposes, we should protect our privacy. What is at stake here is freedom and agency in the area of personal life and experience. We should know who decides and, moreover, why we should give someone the right to decide for us? The use of modern technologies can significantly improve the quality of our lives. However, this kind of improvement must not lead to a violation of those values that

³¹ See Ajay Ar g w a l, Joshua G a n s, and Avi G o l d f a r b, *Predication Machines: The Simple Economics of Artificial Intelligence* (Harvard: Business Review Press, 2018), 13–15.

define the horizon of self-identification of interpersonal relationships, that is, the dignity of the human person and respect for others and their privacy.³²

In the conclusion of his analysis of the impact of AI on human life, Žižek is more pessimistic than Zuboff. He predicts that the combination of the human mind with the collective artificial intelligence will deprive us of the unconscious and will make us unable to hide individual motivation, which will ultimately affect our sense of identity and our moral consciousness. The operation of AI is based on the assumption of transparency; its strength consists in the ability to accumulate, transmit and use the information that is acquired by many devices.³³ Then AI obtains new data, which it subsequently processes in order to find solutions and adapt them to the needs and circumstances. By designing a complex system of human action, AI can influence human motivation. Over time, AI that imitates human intelligence will become transparent to the collective intelligence of institutions. The managers of these institutions will believe that hiding information and cheating will be treated as a crime against the system. This pessimistic scenario does not have to come true; however, it is an important warning as to the ideas concerning the managing of institutions based on the AI model.

POWER OVER AI

There is also this solution: the creators of AI do not want to build conscious machines, but only intelligent machines, ones whose job is to complement human computing skills and thus create a human-friendly world. Here, however, the problem arises whether AI will be able to understand the human world of values and the specific nature of the moral obligations that result from those values. Will artifacts equipped with artificial intelligence be able to recognize the world of human values in their complex nature and will they be able to read the principles that govern the process of making difficult decisions by man? We still do not have satisfactory answers to these questions, even though this issue has been taken up by many researchers, including Eliezer Yudkowsky³⁴ and Bostrom³⁵. When it comes to the knowledge of values, an important role is played by axiological intuition, a distinctly human ability and one which machines lack. At this point, however, another question comes into view, whether

³² See Zuboff, *The Age of Surveillance Capitalism*, 323–5.

³³ See Žižek, *Hegel in a Wired Brain*, 190n.

³⁴ Eliezer Yudkowsky, *Complex Value Systems are Required to Realize Valuable Futures* (San Francisco: Machine Intelligence Research Institute, 2011), <https://intelligence.org/files/ComplexValues.pdf>.

³⁵ Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies*.

it is possible to write down the human structure of values and express it in the form of an algorithm, which can then be inscribed in the operational structure of an intelligent machine.

An advanced-level, super-intelligent AI can perform complex computational operations involving the collecting and segregating of data, while not being aware of its distinctiveness.³⁶ It seems reasonable to argue, however, that an entity cannot be the bearer of responsibility without having moral awareness. To solve complex axiological and moral dilemmas, a machine, besides intelligence, must also have consciousness. A conscious machine, capable of recognizing values and solving moral dilemmas, would need to have a will capable of choosing and acting independently of humans.³⁷ Autonomous machines, fully independent of the man controlling their operation, seem to be—from the perspective of the designers—an undesirable coincidence, unless of course this is also an unintended result of AI techno-evolution, one that we cannot control. This kind of operation, independent of the constructor and the user, is treated as a design error. The argument from the “designer’s unintentional error” has had its reflections in popular culture, in narratives where autonomous robots want to take control over people, who are “less” intelligent.

The creation of an artificial intelligence that imitates the human world of values has also a negative side to it. Apart from positive values, axiology distinguishes negative values, which result in the desire for destruction, death, falsehood, and the creation of distorted (demonic) images of the sacred. Consequently, this leads to behaviors that we consider morally wrong, among them the propensities to be aggressive, to cheat, and to treat other people instrumentally. We cannot assume that man represents the highest level of consciousness and moral competence. Inscribing the human world of values into an intelligent machine can prove problematic. This is the rationale behind the building of “ethical robots,” namely, that the super-intelligent machines thus created will be devoid of human flaws.³⁸ However, this means that this type of ethics becomes a utopian AI construct equipped with an “angelic” set of qualities (such as kindness, forbearance, the ability to cooperate) focused on the fostering of community values while devoid of human “demonic” tendencies. In other words, such projects dehumanize machines and make them into entities which are “artificial” in another sense of the word.

Such a project was created by Yudkowsky, who presented the development of intelligent machines the operation of which is based on positive va-

³⁶ Y u d k o w s k y, *Complex Value Systems are Required to Realize Valuable Futures*, 38n.

³⁷ See Susan S c h n e i d e r, *Artificial You: AI and the Future of Your Mind* (Princeton and Oxford: Princeton University Press, 2019), 16n.

³⁸ See Y u d k o w s k y, *Complex Value Systems are Required to Realize Valuable Futures*, 40n.

lues. To realize this purpose, he used the “semantics of external reference,” which demonstrates that an increase in a machine’s knowledge about values such as kindness will in due course result in an increase in actual kindness in that machine’s operation. In the case of AI, this is the knowledge entered by programmers into the IT system. This solution does not work in the case of human beings in that it does not take into account the factor of free will and the situational dynamics in which the subject who is making the moral decision finds themselves. In fact, it is a machine variant of ethical intellectualism. Intelligent machines, like humans, may know the rules and yet act without conforming to them. Following Aristotle, it is necessary to distinguish in this case between machine *techne* knowledge and machine *praxis* knowledge. The former is responsible for collecting and segregating data, while the latter for choosing and action.

What would the phronetic knowledge of machines consist in? Yudkowsky introduces the formula of “semantics of causal significance” in this case. It assumes that AI should not do exactly what programmers have written into it, but something similar, something that results from a certain skill in solving difficult situations. Developers are not able to take into account all circumstances, so it should be assumed that the AI will be equipped with the ability to modify decisions. This means that the solution proposed by AI does not have to suit us. As the source principle of AI normativity, Yudkowsky adopts: “do the right thing,” which is based on the principle of reflective equilibrium. This is one variant of the Greek rule of the ethics of moderation, introduced by Aristotle into ethical thought.³⁹ Here, however, the problem arises whether the reflective balance of the man being and that of AI are based on the same principles. The premise of the ethics of moderation is to have human experiences resulting from corporeality and communion with other people (ethics of friendship). For AI, experience will only be information inscribed in the system, not an individual or personal experience. Thus, the thesis that a machine can “think and act like a human” is merely an approximation.

The problem emerges of how to prevent AI from wanting to pursue its own goals, ones which would ultimately turn out to be harmful to humans (e.g., the production of a gigantic number of paper clips, which from our point of view means waste). Bostrom is considering a situation when we want to increase the cognitive competence of the system, but we are afraid that this means an increase in its powers to an extent that would ultimately distort the motives that should govern the AI system. Bostrom proposes that the system be divided into a hierarchical structure made of subsystems. Then subsystems with some intelligence potential will monitor the performance of subsystems with fewer

³⁹ See *ibidem*, 43n.

capabilities. The goal is to prevent “thinking” AI subsystems from strategically hiding information or wanting to seize power over the entire system. This they may do, for example, in pursuit of the principle of eliminating the weakest and the least intelligent elements in the system, e.g., humans. In terms of intellectual competence, at the bottom of this hierarchy and at the same time at the very top in the hierarchy of power, one should place a slow and relatively less intelligent superior, that is, man.

One may ask whether such inverted meritocracy guarantees the security and stability of the entire system.⁴⁰ However, there seems to be no certainty that the system will always be safe for us humans. Systems need a hierarchical and multi-level surveillance system in which the number of “workers” is proportional to the number of guards or supervisors, where each of the guards is supervised by a senior guard. Such a system may be stable, but it is based on totalitarian principles of control. This whole structure includes many super-intelligent operators whose actions are controlled by a small group of people with intellectual competence inferior to machines with AI. The question is whether we will remain vigilant here, characteristic of the human way of exercising power; whether we will be ready for the situation when AI moves away from the principles of operation that have been devised for it. As technology progresses, we will entrust it with more and more tasks, which means an increase in trust. At the same time, we will release ourselves from making the unnecessary effort of performing activities that an intelligent machine performs much better than we can. This lack of vigilance could end in a “rebellion” of machines against humans.

The question arises as to how to act when the super-intelligent subsystems decide to choose the “wrong” path and pursue their own goals. Given the history of human societies, this scenario is highly probable, which is why it is not a good idea to ascribe human tendencies to machines. According to Bostrom, this threat can be eliminated by the introduction of programming principles characteristic of intelligent systems into interpersonal relations. So far, our understanding of human behavior has been limited, which is why we are finding it difficult to comprehend the behavior of entities which are not human (AI or Singularities resulting from the connection of the human mind with AI). Intelligent systems can demonstrate the ability to coordinate their activities, with little communication with humans. This could ultimately undermine the ability to control AI and lead to the disintegration of the entire system and the collapse of the institutional order, no matter how safe it may have seemed.⁴¹

⁴⁰ See Bostrom, *Superintelligence: Paths, Dangers, Strategies*, 198n.

⁴¹ See *ibidem*, 205–7.

*

It should be said that further work on AI and its connection with human functioning is inevitable, regardless of whether we see this issue in terms of threats or in terms of new opportunities opening before mankind. This progress is associated with responsibility for the development: there is a need to define ethical and legal limits for the creation of AI. In such situations, we refer to the idea of well-being, while keeping an eye on the two levels of its meaning: the well-being of humanity and the well-being of individuals. We are interested in implementing the postulate of a “better” life at the economic, political, medical, and scientific levels. However, agreeing to improve the quality of life by AI may become a threat to the “good” of humanity. It remains an open question how to create satisfactory rules to regulate the “reflective balance” in the relationship between humans and AI. Teaching human values to a machine seems as yet to be an impossible task, even for machines that surpass humans in intelligence. Even if it were possible to do this, questions arise: What values would we like to enter into AI? Should it really create an orderly and hierarchical system? The solution proposed by Bostrom, i.e., to convey values by designing institutions where man and AI are integrated, is very dangerous. An institutional intelligent value system will not recognize the value of the distinctiveness and independence of individual human beings. Rather, it will focus on the value of teamwork and collective action. This, however, will reconstruct the operation model of a totalitarian institution, as described by Zuboff, and will make possible the pathology of the collective intelligence, as analyzed by Žižek. Due to the exceptionally dynamic development of new technologies, the relationship between man and AI is currently one of the most important issues to ponder. Therefore, this research should be continued.

BIBLIOGRAPHY / BIBLIOGRAFIA

- Argwal, Ajay, Joshua Gans, and Avi Goldfarb. *Prediction Machines: The Simple Economics of Artificial Intelligence*. Harvard: Harvard Business Review Press, 2018.
- Bacon, Francis. *Selected Writings of Sir Francis Bacon*. New York: Franklin Library, 1982.
- . *New Atlantis*. Cambridge: Cambridge University Press, 2013.
- Bostrom, Nick. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press, 2014.
- Cambridge Dictionary, Cambridge University Press & Assessment, s.v. “Domestication.” <https://dictionary.cambridge.org/pl/dictionary/english/domestication>.

- Campanella, Tommaso. *Civitas Solis*. Translated by Daniel J. Donno. Berkeley and Los Angeles: University of California Press, 1981.
- Dusek, Val. *Philosophy of Technology: An Introduction*. Oxford: Blackwell Publishing Ltd., 2006.
- Fleischer, Michael. *Kapitalizm i jego sztuczna inteligencja*. Wrocław and Kraków: Wydawnictwo Libron, 2022.
- Hegel, Georg Wilhelm Friedrich. *Fenomenologia ducha*. Translated by Światosław Florian Nowicki. (Warszawa: Fundacja Aletheia, 2002).
- Kamińska, Joanna. *Nowe wspaniałe światy: Współczesne projekty doskonałego społeczeństwa*. Kraków: Nomos, 2012.
- Klain, Stephen J. "What is Technology?" *Bulletin of Science, Technology & Society*, no. 5 (1985): 215–8.
- Kurzweil, Ray. *The Singularity Is Near: When Humans Transcend Biology* (London: Penguin Publishing Group, 2006).
- Manuel, Frank E., and Fritzie P. Manuel. *Utopian Thought in Western World*. Cambridge: Belknap Press, 1979.
- More, Thomas. *Utopia*. Translated by Clarence H. Miller. New Haven and Yale: Yale University Press, 2001.
- Pacey, Arnold. *The Culture of Technology*. Oxford: Blackwell Publishing Ltd., 1983. <https://www.fulcrum.org/epubs/3x816q10j?locale=en#page=18>.
- Schneider, Susan. *Artificial You: AI and the Future of Your Mind*. Princeton and Oxford: Princeton University Press, 2019.
- Skinner, Burrhus F. *About Behaviorism*. New York: Vintage Books, 1976.
- . *Science and Human Behaviour*. New York: The Free Press, 1965.
- . *Walden Two: With a New Preface by Author*. Cambridge: Hackett Publishing Company, 2005.
- Stephenson, David. *Big Data Demystified: How to Use Big Data, Data Science and AI to Make Better Business Decisions and Gain Competitive Advantage*. London: Pearson Education, 2011.
- Szacki, Jerzy. *Spotkania z utopią*. Warszawa: Państwowy Instytut Wydawniczy, 1980.
- Walicki, Andrzej. *Marksizm i skok do królestwa wolności. Dzieje komunistycznej utopii*. Warszawa: Wydawnictwo Naukowe PWN, 1996.
- Wikipedia, s.v. "Technological Utopianism." https://en.wikipedia.org/wiki/Technological_utopianism.
- Yudkowsky, Eliezer. *Complex Value Systems are Required to Realize Valuable Futures*. San Francisco: Machine Intelligence Research Institute, 2011. <https://intelligence.org/files/ComplexValues.pdf>.
- Zuboff, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: Perseus Books LLC, 2019.
- Žižek, Slavoj. *Hegel in a Wired Brain*. London: Bloomsbury Publishing, 2020.

ABSTRACT / ABSTRAKT

Mariusz WOJEWODA, *Artificial Intelligence as a Social Utopia*

DOI 10.12887/36-2023-3-143-08

In the article, the author is concerned with the problem of artificial intelligence as a factor that affects the functioning of social institutions. AI is understood as the logic of machine thinking, which constitutes a foundation for the introduction of optimal operating principles for the smooth functioning of the system. This provides the context for the author's analysis of the logic of power in two aspects: (1) AI as a model for exercising control over members of institutions; (2) the possibility of exercising control over AI by humans. In both cases, we are dealing with a variant of the social utopia, the assumption being that such control brings the desired effects, namely, an increase in the efficiency of the system involving people and machines. In both cases, we deal with threats to freedom of choice. To demonstrate the complexity of the issue, the author analyzes the concepts of Shoshana Zuboff, Slavoj Žižek, Nick Bostrom, and Elizer Yudkowsky. These theorists point to risks and opportunities that come with the use of AI. The AI-based utopia offers us a digital paradise based on security and unlimited access to information, while at the same time taking away the right to freely decide for ourselves. This process is so subtle that the fascination with the use of AI tends to ignore the scale of the threat. The author is seeking to examine the validity of dystopian objections against AI.

Keywords: techno-utopia, artificial intelligence, AI ethics, Shoshana Zuboff, Slavoj Žižek, Nick Bostrom, Elizer Yudkowsky

Contact: Instytut Filozofii Nauki, Wydział Humanistyczny, Uniwersytet Śląski, ul. Bankowa 11, 40-007 Katowice, Poland

E-mail: mariusz.wojewoda@us.edu.pl

Phone: +48 32 3591709

<https://us.edu.pl/instytut/ifil/osoby/mariusz-wojewoda/>

Mariusz WOJEWODA – Sztuczna inteligencja jako utopia społeczna

DOI 10.12887/36-2023-3-143-08

Celem artykułu była analiza problemu sztucznej inteligencji, która wpływa na funkcjonowanie instytucji społecznych. AI rozumiana jest jako logika myślenia maszynowego, która zakłada wprowadzenie optymalnych zasad działania w celu sprawnego funkcjonowania systemu. W tym kontekście autor analizuje logikę władzy w dwóch aspektach: (1) AI jako modelu sprawowania kontroli nad członkami instytucji, (2) możliwości sprawowania kontroli nad AI przez człowieka. W obu przypadkach mamy do czynienia z pewnym wariantem utopii społecznej, w której zakłada się, że taka kontrola przynosi pożądane efekty dla podniesienia sprawności działania systemu – ludzi i maszyn. W obu

wypadkach musimy uporać się z zagrożeniami dotyczącymi wolności wyboru. Aby ukazać złożoność tego zagadnienia, autor analizuje koncepcje Shoshany Zuboff, Slavoj Žižka, Nicka Bostroma i Elizera Yudkowsky'ego. Teoretycy ci wskazują na zagrożenia i możliwości, jakie wiążą się z wykorzystaniem AI. Utopia oparta na AI oferuje nam cyfrowy raj, zapewniający bezpieczeństwo i nieograniczony dostęp do informacji, a jednocześnie odbiera prawo do wolnego decydowania o sobie. Działanie to jest na tyle subtelne, że fascynacja związana z możliwościami wykorzystania AI prowadzi do niedostrzegania skali zagrożenia. Autor artykułu chce sprawdzić słuszność dystopijnych zarzutów stawianych AI.

Słowa kluczowe: techno-utopia, sztuczna inteligencja, etyka AI, Shoshana Zuboff, Slavoj Žižek, Nick Bostrom, Elizer Yudkowsky

Kontakt: Instytut Filozofii Nauki, Wydział Humanistyczny, Uniwersytet Śląski,
ul. Bankowa 11, 40-007 Katowice

E-mail: mariusz.wojewoda@us.edu.pl

Tel. 32 3591709

<https://us.edu.pl/instytut/ifil/osoby/mariusz-wojewoda/>

Grażyna OSIKA

KOMUNIKACJA CZŁOWIEKA ZE SZTUCZNĄ INTELIGENCJĄ Warunki brzegowe

Wśród wielu aspektów, jakie musimy wziąć pod uwagę, istotne wydają się te czynniki, które będą przesądzały o rodzaju relacji zbudowanych z maszynami jako nowymi agentami „społecznymi”. Dotyczy to szczególnie tej najbardziej wyrażonej formy określanej mianem sztucznej inteligencji (SI). Rozumienie tych relacji wiąże się z doprecyzowaniem, jakim bytem są dla nas maszyny. W filozofii techniki doczekaliśmy się kilku podejść, od najbardziej intuicyjnego, jakim jest instrumentalizm, przez substancywizm, aż po postfenomenalizm, równoważący oba stanowiska.

W *The World Manufacturing Forum Report: Skills for the Future of Manufacturing* w roku 2019 na podstawie przeprowadzonych badań przewidywano, że wśród dziesięciu kluczowych umiejętności przyszłości znajdzie się skuteczna komunikacja między ludźmi, systemami IT i AI¹. Pandemia COVID-19 i obecna sytuacja na świecie pokazała, jak trafne były te przewidywania, ale także, że nie straciły one na aktualności, wręcz przeciwnie – trend ten się umocnił. Informacji tej nie powinniśmy traktować jedynie jako „branżowej” prognozy, stoją za nią bowiem implikacje dotyczące kształtu, jaki planujemy nadać przyszłemu społeczeństwu, co będzie stanowiło clou naszej aktywności, a także będzie nas definiowało jako ludzi.

W rozwijanych koncepcjach często przewija się pojęcie Społeczeństwa 5.0², w którym „widzimy” siebie jako żyjących wśród maszyn, jednak w tej wizji urządzenia nie są już prostymi narzędziami przedłużającymi siłę naszych

¹ Por. *The World Manufacturing Forum Report. Skills for the Future of Manufacturing*, <https://worldmanufacturing.org/wp-content/uploads/WorldManufacturingFoundation2019-Report.pdf>, s. 32-34. M. Breque, L. De Nul, A. Petridis, European Commission, Directorate-General for Research and Innovation, *Industry 5.0: Towards a Sustainable, Human-centric and Resilient European Industry*, Publications Office, 2021, <https://data.europa.eu/doi/10.2777/308407>.

² Zob. S. Arsovski, *Quality of Life and Society 5.0*, International Quality Conference 13 IQC Quality Research 2019, http://www.cqm.rs/2019/papers_iqc/81.pdf; M. G l a d d e n, *Who Will Be the Members of Society 5.0? Towards an Anthropology of Technologically Posthumanized Future Societies*, „Social Science” 148(2019) nr 8(5), s. 1-39; A. Deguchi, H. Chiaki, M. Hideyuki, T. Shigeyuki, N. Taku, O. Kohei, *What is Society 5.0?*, w: Hitachi-UTokyo Laboratory, *Society 5.0: A People-centric Super-smart Society*, Springer Open, Singapore 2020.

mięśni, upraszczającymi działanie lub czyniącymi je w ogóle możliwym³, tym razem chodzi o ekstensję tego, co w potocznym mniemaniu uznajemy za jeden z najbardziej znaczących wyróżników naszego gatunku, czyli myślenie. Erik Brynjolfsson i Andrew McAfee uważają, że dopiero ten typ maszyn stanie się prawdziwym przełomem w historii ludzkości⁴. Czy mamy podstawy powątpiewać w ich prognozy? Codziennosc wskazuje na retoryczny charakter tego pytania, albowiem – jak to kolokwialnie określamy – „gołym okiem widać”, że kolejne intensyfikacje technologiczne, takie jak sztuczna inteligencja (ang. artificial intelligence, AI), opierająca się na cyfrowych rozwiązaniach założycielskich (komputerze i Internecie)⁵, wnikają w każdy aspekt naszego życia, jednocześnie go przeobrażając. Szeroko rozumiane funkcjonowanie społeczne, w tym nasza aktywność zawodowa, są i będą uwikłane technologicznie – jest to już faktem, stąd potrzeba adaptowania się do nowej sytuacji. Wśród wielu aspektów, jakie musimy wziąć pod uwagę, istotne wydają się te czynniki, które będą przesądzały o rodzaju relacji zbudowanych z maszynami jako nowymi agentami „społecznymi”. Dotyczy to szczególnie tej najbardziej wyrafinowanej formy określanej mianem sztucznej inteligencji (SI). Rozumienie tych relacji wiąże się z ustaleniem czegoś jeszcze bardziej podstawowego, a mianowicie doprecyzowaniem, jakim bytem są dla nas maszyny. W filozofii techniki doczekaliśmy się kilku podejść, od najbardziej intuicyjnego, jakim jest instrumentalizm, wskazującego na nadrzędną pozycję użytkownika wobec „narzędzia”, przez substancywizm, w ramach którego maszyny (technika) pozostają w znacznym stopniu autonomiczne wobec człowieka, aż po postfenomenalizm, równoważący oba stanowiska. Co prawda każde narzędzie wpływa na określoną formę jego użycia, jak przyjmuje się w substancywizmie, ale ostatecznie to człowiek w trakcie pracy ustala, w jaki sposób z danego narzędzia będzie korzystał – dochodzi zatem do mediacji między człowiekiem a maszyną⁶. I tak według założeń instrumentalizmu „komuniko-

³ Por. G. O s i k a, *Czekając na osobliwości – o modelach interpretacji techniki*, „Filo-Sofija” 17(2017) nr 39(1), s. 68.

⁴ Por. E. B r y n j o l f s s o n, A. M c A f e e, *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*, W.W. Norton & Company, New York, London 2014, s.10.

⁵ Por. K. Ś l e d z i e w s k a, R. W ł o c h, *Gospodarka cyfrowa. Jak nowe technologie zmieniają świat*, Wydawnictwa Uniwersytetu Warszawskiego, Warszawa 2020, s. 79.

⁶ Zob. *The Oxford Handbook of Philosophy of Technology*, red. S. Vallor, Oxford University Press, New York 2022; A. F r e e n b e r g, *Transforming Technology: A Critical Theory Revisited*, Oxford University Press, New York 2002; t e n ż e, *Critical Theory of Technology: An Overview*, „Tailoring Biotechnologies” 1(2005), s. 47-64; A. B o r g m a n n, *Technology and the Character of Contemporary Life: A Philosophical Inquiry*, University of Chicago Press, Chicago 1984; T. M i c z k a, *Człowiek techniczny, „człowiek rozszerzony”, „człowiek 3D”. Próba definicji*, „Filo-Sofija” 17(2017) nr 39(1), s. 37-47.

wać się” z narzędziem to wydawać instrukcje lub odbierać je, w przypadku substancywizmu jest mowa raczej o tak zwanej interakcji zamkniętej⁷, czyli takiej, która może dokonywać się jedynie w ramach interfejsu maszyny. Dopiero postfenomenalizm wraz z kategorią mediacji otwiera się na rozumienie komunikacji jako wzajemnego oddziaływania. Widać zatem, że komunikacja między SI a człowiekiem nie jest jedynie pochodną możliwości technicznych systemu komputerowego, ale także sposobu, w jaki system ten jesteśmy skłoni potraktować – jako „nie-ludzkie”⁸ agenty współdziałania.

Nabywanie umiejętności w jakimkolwiek zakresie zawsze wymaga elementarnej wiedzy dotyczącej mechanizmów związanych z tymi umiejętnościami. Jest to tym bardziej istotne, kiedy planujemy włączać je do działań edukacyjnych, pomagających wykształcić kompetencje „dla przyszłości” w duchu założeń Społeczeństwa 5.0. Z tej perspektywy ważne jest przebadanie warunków kluczowych dla zachodzenia komunikacji człowieka z SI. W niniejszych rozważaniach podejmuje się próbę odpowiedzi na dwa pytania badawcze. Pierwsze pytanie: jakie warunki brzegowe⁹ musi spełnić komunikacja? Odpowiedź na nie stanowi część prac regulujących definiowanie pojęcia „komunikacja”. Pozwoli to ustalić warunki modelowe dla zachodzenia tego procesu. Drugie pytanie to: jakie warunki komunikacyjne spełnia SI jako program komputerowy wykorzystujący zaawansowane modele statystyczne do rozwiązywania zadań? W tym przypadku chodzi o diagnozę, jak dalece wskazywane wcześniej warunki komunikacji leżą w zakresie wykonalności systemów SI, czyli na jaki poziom „skomunikowania” człowieka z maszyną możemy obecnie liczyć. Prace te mogą być pomocne w rozstrzygnięciu, w jakim stopniu mediacyjne podejście do relacji człowiek–maszyna jest w ogóle uzasadnione oraz jak modelować umiejętności komunikacyjne człowieka z SI.

W badaniach zostanie wykorzystana analiza konceptualna¹⁰, pozwalająca na podstawie obecnego stanu wiedzy rozwinąć teoretyczny namysł, który w toku prac empirycznych będzie mógł następnie zostać poddany weryfikacji, a więc na tym etapie badań metoda ta wydaje się w pełni uzasadniona.

Ze względu na to, że artykuł nie ma charakteru przeglądowego, nie wydaje się konieczne odtwarzanie historii rozwoju pojęcia komunikacji, chodzi

⁷ Por. L. M a n o v i c h, *Język nowych mediów*, tłum. P. Cypryański, Wydawnictwa Akademickie i Profesjonalne, Warszawa 2006, s. 128-135.

⁸ Por. A. P r z e g a l i Ń s k a, D. J e m i e l n i a k, *Społeczeństwo współpracy*, Wydawnictwo Naukowe Scholar, Warszawa 2020, s. 137.

⁹ Warunki brzegowe to „coś, co jest absolutnie wymagane w danej sytuacji” (hasło: „Warunek brzegowy”, w: *Wielki słownik języka polskiego*, https://wsjp.pl/haslo/do_druku/97871/warunek-brzegowy).

¹⁰ Por. E. J a k k o l a, *Designing Conceptual Articles: Four Approaches*, „AMS Review” 10(2020), s. 18-26.

raczej o uchwycenie kluczowych momentów umożliwiających wyodrębnienie warunków brzegowych dla zachodzenia tego procesu.

KOMUNIKACJA – WARUNKI BRZEGOWE

Zaproponowana w niniejszym artykule metoda badawcza wymaga wyboru adekwatnych narzędzi teoretycznych, a tego typu decyzje zawsze w jakimś stopniu są obciążone arbitralnością wynikającą z życiorysu intelektualnego autora. Jednakże niezależnie od świadomości tych zagrożeń, równie zawodne jest założenie, że należy brać pod uwagę wszystkie tradycje teoretyczne. Dylemat ten jest szczególnie uzasadniony w odniesieniu do teorii komunikacji, na co zwracali uwagę badacze od samego początku konstituowania się dyscypliny¹¹. Główna trudność polega na tym, że „proponowane w ramach konkretnych teorii definicje stanowią element tych teorii, niejako «pracują» na nie. Różne definicje komunikacji pełnią odmienne funkcje i są przez teoretyków wykorzystywane w różny sposób”¹², na przykład czasami komunikacja rozpatrywana jest nie ze względu na nią samą, na jej cechy, ale jako element szeroko rozumianych systemów¹³.

W niniejszych rozważaniach klucz doboru stanowi odniesienie się do tych teorii, które za punkt wyjścia przyjmują dostrzeżenie określonego fenomenu cechującego komunikację i jednocześnie są obecnie niekwestionowaną składową wiedzy o procesie komunikacji. Są to odkrycia wyłaniające się w toku obserwacji praktyk komunikacyjnych, dlatego analizie nie grozi błąd nieadekwatności wykorzystanych modeli, odwołujemy się bowiem bezpośrednio do podstawowych aspektów doświadczenia komunikacyjnego.

Zaproponowane podejście należy traktować jako wiedzę skumulowaną, uzyskaną dzięki szczegółowym spostrzeżeniom poszczególnych badaczy.

¹¹ Zob. F. D a n c e, *Human Communication Theory: Original Essays*, Holt, Rinehart & Winston, New York 1967; R.T. C r a i g, *Communication Theory as a Field*, „Communication Theory” 9(1999) nr 2, s. 119-161; K. M i l l e r, *Communication Theories: Perspectives. Processes and Context*, McGraw Hill, Boston 2005; S.W. L i t t l e j o h n, K.A. F o s s, *Theories of Human Communication*, Waveland Press, Long Grove 2011, s. 4n.; M. W e n d l a n d, *Konstruktywizm komunikacyjny*, Wydawnictwo Naukowe UAM, Poznań 2011, s. 189-199; E. K u l c z y k i, *Teoretyzowanie komunikacji*, Wydawnictwo Naukowe UAM, Poznań 2012; J. P l e s z c z y Ń s k i, *Epistemologia komunikacji medialnej. Perspektywa ewolucyjna*, Wydawnictwo UMCS, Lublin 2013, s. 40-44.

¹² G. O s i k a, *Tożsamość osobowa w epoce cyfrowych technologii komunikacyjnych*, Universitas, Kraków 2016, s. 157.

¹³ Doskonałym przykładem takiego podejścia jest fragment tekstu Michaela Fleischera: „Komunikacja nie jest niczym innym, jak stosującym znaki i sterowanym przez nie mechanizmem orientacyjnym i negocyjacyjnym służącym do wytworzenia i zabezpieczenia systemu społecznego” (M. F l e i s c h e r, *Zarys ogólnej teorii komunikacji*, w: *Radykalny konstruktywizm. Antologia*, red. B. Balicki, D. Lewiński, B. Ryż, E. Szczerbuk, Centrum im. Willy Brandta, Wrocław 2010, s. 345; zob. N. L u h m a n n, *Systemy społeczne. Zarys ogólnej teorii*, tłum. M. Kaczmarczyk, Nomos, Kraków 2012).

W celu łatwiejszego „poukładania myśli” zostanie wykorzystane ujęcie tabelaryczne (por. tabela 1), w którym uwzględniono takie wymiary, jak: (1) ustalenie badawcze – zawierające informacje, kiedy, kto oraz na co zwrócił uwagę; (2) istota ustalenia – czyli waga ustalenia, z punktu widzenia rozumienia komunikacji; (3) warunek zachodzenia komunikacji – konkretny warunek, jaki musi zostać wzięty pod uwagę; (4) składowe definicji wynikające z określonego ustalenia.

Tabela 1. Warunki brzegowe procesu komunikacji

	Ustalenia badawcze	Istota ustaleń	Warunek zachodzenia komunikacji	Składowe definicji wynikające z określonego ustalenia
1.	1948 – Harold Lasswell: pytania kluczowe z punktu widzenia procesu komunikacji: kto mówi, co mówi, jakim kanałem, do kogo, po co ¹⁴ .	Ustalenie elementów składowych procesu komunikacji.	Na proces komunikacyjny składają się: nadawca, przekaz, kanał, odbiorca oraz cel, dla którego się komunikujemy.	Znamy elementy konieczne do zachodzenia procesu komunikacji.
2.	1952 – Friedrich Waismann: otwartość znaczeniowa (ang. open texture) ¹⁵ .	Akceptacja możliwości nieostrości denotacyjnej nośników znaczenia, wynikających ze zmian rzeczywistości.	Każdy „znak” ma częściowo niestabilne znaczenie.	Wiemy, że ostateczne znaczenie „wydobywa” się „w użyciu” jego nośnika.
3.	1954 – Yehoshua Bar-Hillel: uznanie roli kontekstu w rozpoznawaniu znaczenia wypowiedzi ¹⁶ .	Ustalenie kontekstowości procesu komunikacji.	Konieczna znajomość kontekstu dla ustalenia znaczenia przekazu.	Wiemy, że musimy znać kontekst, aby doprecyzować znaczenie przekazu.
4.	Wprowadzenie kategorii sprzężenia zwrotnego do procesu komunikacji przypisuje się tradycji cybernetycznej, to jest matematycznemu modelowi Shannona i Weavera ¹⁷ (1948/1964). Chociaż	Uznanie roli sprzężenia zwrotnego w procesie komunikacji.	Między uczestnikami komunikacji dokonuje się sprzężenie zwrotne pozwalające weryfikować przepływ informacji.	Komunikacja jest interakcją.

¹⁴ Por. H.D. L a s s w e l l, *The Structure and Function of Communication in Society*, w: *The Communication of Ideas*, red. L. Bryson, Harper and Row, New York 1948, s. 37-51.

¹⁵ Por. F. W a i s m a n n, *Verifiability*, „Journal of Symbolic Logic” 12(1947) nr 3, s. 101.

¹⁶ Zob. Y. B a r - H i l l e l, *Indexical Expressions*, „Mind, New Series” 251(1954), s. 359-379.

¹⁷ Zob. C.E. S h a n n o n, W. W e a v e r, *The Mathematical Theory of Communication*, The University of Illinois Press, Urbana 1964.

	oni sami nie zawarli go w swoim schemacie, jednak większość badaczy uznaje, że w obrębie tego podejścia doszło do jej wyłonienia bez konkretnego wskazywania, kto tego dokonał.			
5.	1962 – John L. Austin: uzmysłowienie performatywnej funkcji komunikacji ¹⁸ .	Ustalenie, że mówienie (komunikacja) jest równoznaczne z wykonywaniem czynności.	Ustalenie czynności, jakie wykonujemy, się komunikując.	Wiemy, że komunikacja jest działaniem.
6.	1962-1967 – John L. Austin, John Searle: akt komunikacyjny składa się z aktu wypowiedzenia (lokucja) i wykonania czynności (illokucja) ¹⁹ .	Rozdzielenie, aktu „wypowiedzenia” od właściwego wykonania czynności komunikacyjnej.	Warunek „wypowiedzenia” nie jest wystarczający do zachodzenia komunikacji.	Wiemy, że „mówić” nie znaczy jednocześnie „się komunikować”.
7.	1963 – Roman Jakobson: określenie funkcji językowych – ekspresyjnej, referencyjnej, poetycznej, konatywnej, fatycznej, metajęzykowej ²⁰ .	Rozpoznanie możliwych zastosowań komunikacji powiązane z tym, do czego komunikacja się odnosi.	Realizowany cel komunikacyjny – komunikacja się tematyzuje.	Wiemy, że komunikacja jest procesem celowym – komunikacja jest „o czymś” i „po coś” i oba wymiary są od siebie zależne.
8.	1967 – Paul Herbert Grice: ustalenie warunków zachodzenia skutecznej konwersacji ²¹ .	Ustalenie istnienia tak zwanych implikatur, czyli „wnioskowań asemantycznych”, kiedy odbiorca „nadpisuje” właściwy sens, mając wiedzę na temat dosłownego znaczenia wypowiedzianych słów oraz kontekstów, w których do komunikacji dochodzi.	Konieczność uwzględnienia interpretacji odbiorcy.	Wiemy, że odbiorca jest aktywny w kształtowaniu znaczenia przekazu.

¹⁸ Zob. J.L. A u s t i n, *How to do Things With Words*, Oxford University Press, London 1962.

¹⁹ Zob. tamże; J.R. S e a r l e, *Speech Acts: An Essay in the Philosophy of Language*, Cambridge University Press, Cambridge, New York, Melbourne, Madrid 1967; G. O s i k a, *The Communicative Acts as Action*, „Lingua as Communitas” 18(2008), s. 35-44.

²⁰ Por. R. J a c o b s o n, *Essais de linguistique generale*, Éditions de Minuit, Paris 1963; G. O s i k a, *Procesy i akty komunikacyjne*, Universitas, Kraków 2011, s. 70n.

²¹ Zob. P.H. G r i c e, *Logic and Conversation*, w: *Studies in the Way of Words*, Harvard University Press, Cambridge, Massachusetts, London 1967, s. 44-58.

9.	1981 – Everett M. Rogers, Lawrence D. Kincaid: konwergencyjny model komunikacji ²² .	Włączenie do procesu komunikacji całego zaplecza poznawczego jako elementu konstytutywnego.	Uwspólnione działanie na poziomie percepcji, interpretacji, rozumienia.	Wiemy, że komunikacja jest współdziałaniem, które jest efektem konwergencji wielu czynników: percepcji, interpretacji, rozumienia, przekonania.
10.	1987 – Guy Fielding, Peter Hartley: ludzie potrafią podczas komunikacji rekompensować niedobór sygnałów za pomocą innych wskazówek ²³ .	Podczas komunikacji wykazujemy się dużą elastycznością, jeżeli chodzi o dobór wykorzystywanych nośników znaczenia.	Utrata konwencjonalnego nośnika znaczenia nie oznacza zaprzestania komunikacji, zaczynamy inne sygnały traktować jako nośniki, na przykład przy rozmowie przez telefon bardziej znaczące stają się walory lokalizacyjne.	Wiemy, że elastycznie dostosowujemy dobór sygnałów do budowania znaczenia, z których korzystamy podczas komunikacji.
11.	1992 – Byron Reeves i Clifford Nass: opracowanie koncepcji tak zwanych starych mózgów – stosunkowo krótki czas, w procesie naszej ewolucji kontaktu z mediami, czyli technicznymi narzędziami zapośredniczającymi ²⁴ komunikację, powoduje, że ludzkie reakcje na media mają charakter społeczny i naturalny ²⁵ .	Media wykorzystywane w procesie komunikacji traktujemy jak aktorów społecznych.	Przebieg komunikacji wyzwalający z ograniczeń „tu” i „teraz” wymaga użycia medium, które traktujemy jak aktora społecznego.	Wiemy, że komunikację odbywającą się za pośrednictwem urządzeń służących do komunikacji traktujemy tak samo jak tę, która przebiega w sposób bezpośredni, czyli face to face.
12.	2008 – Michael Tomasello: warunkiem komunikacji jest współdzielona uwaga i współdzielona intencjonalność ²⁶ .	Znamy rodzaj zaangażowania konieczny do zachodzenia procesu komunikacji, są to uwaga i motyw.	Komunikacja wymaga podzielania uwagi i intencji.	Wiemy, że elementarnymi warunkami zachodzenia komunikacji są współdzielona uwaga i intencjonalność, dotycząca co najmniej chęci komunikowania się.

Źródło: Opracowanie własne.

²² Zob. E.M. Rogers, L.D. Kincaid, *Communications Networks: Toward a New Paradigm for Research*, Free Press, New York 1981; O s i k a, *Tożsamość osobowa w epoce cyfrowych technologii komunikacyjnych*, s. 157-159.

²³ Por. G. Fielding, P. Hartley, *The Telephone: Neglected Medium*, w: *Studies of Communication*, red. A. Cashdan, M. Jordin, Basil Blackwell, Oxford 1987, s. 110-124.

²⁴ Komunikację zapośredniczoną w niniejszych rozważaniach należy rozumieć jako dokonującą się za pomocą określonego środka technicznego, który „uczestniczy” w przekazie.

²⁵ Zob. B. Reeves, C. Nass, *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Place*, Cambridge University Press, New York 1992.

²⁶ Zob. M. Tomasello, *A Natural History of Human Thinking*, MIT Press, Cambridge 2008.

Uwzględniając ustalenia wynikające z przytaczanych spostrzeżeń poszczególnych badaczy, można pokusić się o sformułowanie opartej na nich definicji komunikacji. Komunikacja jest to dynamiczny proces nakierowany na współdziałanie społeczne, w którym uczestnicy (nadawcy i odbiorcy), na skutek współdzielonej uwagi i intencjonalności oraz ciągłego doprecyzowania znaczenia przekazu, możliwemu dzięki otwartości denotacyjnej, kontekstowi, wnioskowaniom asemantycznym, sprzężeniu zwrotnemu i innym, realizują zakładane cele – formułowane w treści, wykorzystując do tego dostępne, właściwe do nich, nośniki znaczenia. Powyższe dotyczy także komunikacji dokonującej się za pośrednictwem technicznych środków przekazu.

Adekwatnie do opisanej definicji za „sytuację komunikacyjną” uznaje się zespół okoliczności, który towarzyszy inicjowaniu, kontynuowaniu bądź przerywaniu procesu komunikacji. „Sytuacja komunikacyjna” jest elementem modyfikującym dokonującą się komunikacji, ale także sama podlega modyfikacji wskutek przebiegu procesu²⁷. Natomiast wyrażenie „komunikowanie się” należy rozumieć zgodnie z zasadami stosowania zaimka zwrotnego „się”, czyli „że wykonawca czynności sam jej podlega”²⁸; komunikować się to być uczestnikiem procesu komunikacji.

Zarówno rozpoznania poszczególnych badaczy, jak i przytoczone definicje wskazują na elastyczność jako kluczową cechę komunikacji. W ludzkiej komunikacji nie da się uniknąć niestabilności znaczenia wykorzystywanych nośników oraz narzędzi, kontekstowości czy też stałych wzajemnych oddziaływań. To właśnie ta elastyczność pozwala nam aktywnie dostosowywać się do sytuacji w trakcie przebiegu procesu, także wtedy, kiedy komunikacja dokonuje się za pomocą określonego środka przekazu.

Wszelkie obserwacje wskazują na to, że warunkiem elastyczności komunikacyjnej jest świadomość, którą roboczo w charakterze minimalnego jej wymogu możemy rozumieć jako zorganizowaną strukturę układu nerwowego²⁹. „W ludzkim mózgu świadome doświadczenie składa się z wielu składników ujawniających się jednocześnie”³⁰. Informacje zmysłowe zostają przetworzone „w różnych obszarach ludzkiego mózgu, a następnie zintegrowane w spójną całość i (jakość) przekazane do «świadomego ja» (czymkolwiek ono jest) w formie sensownego całościowego doświadczenia”³¹. Integracja ta dokonuje się na bieżąco w odpowiedzi na reakcje dostosowawcze do sytuacji, dlatego zbyt sztywne „zaprogramowa-

²⁷ Por. O s i k a, *Procesy i akty komunikacyjne*, s. 184.

²⁸ Zob. hasło „Zaimek zwrotny”, w: *Słownik języka polskiego PWN*, <https://sjp.pwn.pl/slowniki/zaimek%20zwrotny.html>.

²⁹ Por. P. D a v i e s, *Demon w maszynie. Jak ukryte sieci informacji wyjaśniają tajemnicę życia*, tłum. T. Lanczewski, Copernicus Center Press, Kraków 2020, s. 237.

³⁰ Tamże.

³¹ Tamże.

nie” działania lub zbyt stabilna struktura nośników znaczenia niesie zagrożenie braku relewancji, nieadekwatności, braku porozumienia, czyli wszystkiego tego, czym komunikacja nie jest³². Kiedy myślimy o komunikacji z SI (maszyną), należy pytać o możliwy zakres jej elastyczności komunikacyjnej. Najpierw jednak trzeba zrozumieć z „kim” (czym) ta komunikacja ma się dokonywać.

SZTUCZNA INTELIGENCJA

Początki debaty na temat AI sięgają konferencji w Dartmouth w 1956 roku³³, w której między innymi brali udział tacy badacze, jak: Claude E. Shannon z Bell Labs (teoria informacji), Marvin Minsky z Harvardu (matematyk), Mellon Herbert Simon z Carnegie (ekonomista), John McCarthy z Harvardu (psycholog), George Miller (znany z prac nad ludzką pamięcią) oraz John Nash (matematyk, laureat nagrody Nobla). Termin „sztuczna inteligencja” sformułował wówczas McCarthy, nadając oficjalnie tę nazwę nowemu projektowi inżynierii inteligentnego życia. Założenia inicjatywy były proste: zbadać naturę zdolności poznawczych (myślenia), zaprojektować program odtwarzający te zdolności, a następnie wdroić i przetestować jego działanie na powstających komputerach³⁴. SI miała być nauką „o tworzeniu maszyn zdolnych do wykonywania zadań, które wymagałyby inteligencji, gdyby były wykonywane przez ludzi lub w mniejszym, słabszym sensie, przez zwierzęta”³⁵. Obecnie pod pojęciem SI rozumie się multidyscyplinarną dziedzinę inżynierii, obejmującą robotykę, sieci neuronowe, uczenie maszynowe, A-Life oraz logikę rozmytą. Jako dziedzina badań naukowych korzysta ona z dorobku informatyki (ang. computer science), neurokognitywistyki, biologii, teorii systemów i organizacji, filozofii i neuropsychologii³⁶. Za tymi dość skomplikowanymi terminami w praktyce kryją się pragmatyczne cele. Są to: opracowywanie modeli inteligentnych zachowań oraz budowanie programów zdolnych te zachowania symulować bądź odtwarzać³⁷. Sztuczna inteligencja za-

³² Zob. B. B r o Ź e k, *Granice interpretacji*, Copernicus Center Press, Kraków 2014.

³³ Por. E.L. L a r s o n, *The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do*, The Belknap Press of Harvard University Press, Cambridge, Massachusetts, London 2021, s. 50.

³⁴ Tamże T. W a l s h, *To żyje! Sztuczna inteligencja. Od logicznego fortepianu do zabójczej broni*, Wydawnictwo Naukowe PWN, Warszawa 2018, s. 32n.; M. S u c h a c k a, R. M u s t e r, M. W o j e w o d a, *Human and Machine Creativity: Social and Ethical Aspects of the Development of Artificial Intelligence*, „Creativity Studies” 14(2021) nr 2, s. 434.

³⁵ C.H. H o f f m a n, *Is AI Intelligent? An Assessment of Artificial Intelligence, 70 years after Turing*, „Technology in Society” 68(2022), s. 2.

³⁶ Por. A. P r z e g a l i Ń s k a, P. O k s a n o w i c z, *Sztuczna inteligencja. Nieludzka, arcyłudzka*, Wydawnictwo Znak, Kraków 2020, s. 43.

³⁷ Por. tamże.

tem to „określenie programu komputerowego, który zamiast wykonywać listę dokładnych instrukcji, wykorzystuje zaawansowane modele statystyczne do tego, żeby rozwiązać zadanie. Inteligencja w tym przypadku oznacza, że program na podstawie zbioru danych zawierającego przykłady zadań i poprawne odpowiedzi – sam znajduje występujące między nimi zależności (wzorce)”³⁸. Od samego początku było oczywiste, że prace nad SI nie będą miały charakteru jednorodnego, wyróżniano sztuczną inteligencję tak zwaną słabą (ang. weak artificial intelligence) lub inaczej wąską (ang. narrow artificial intelligence, ANI) i silną (ang. strong artificial intelligence), określaną także mianem pełnej (ang. full) albo uniwersalnej (ang. general artificial intelligence, AGI)³⁹ oraz superinteligencję (osobliwość, ang. singularity)⁴⁰. W słabej (wąskiej) AI chodzi o doskonalenie wykonania danej czynności; do niej zalicza się aplikacje specjalizujące się w jednym obszarze. Klasycznym przykładem jest Deep Blue IBM (ANI). Silny (uniwersalny, pełny) jest to stopień skomplikowania oprogramowania, przy którym aplikacje mają osiągać poziom ludzkiej inteligencji, czyli móc myśleć abstrakcyjnie i logicznie, szybko uczyć się, a także zdobywać wiedzę na podstawie doświadczenia (AGI). Natomiast w przypadku superinteligencji (osobliwości) komputery mają stać się autonomiczne i sprawniejsze intelektualnie od człowieka. Obecnie jesteśmy na poziomie pierwszym, używamy wyspecjalizowanych aplikacji, które stanowią wsparcie w wąskich obszarach działania. Nie jest to jednak równoznaczne z tym, że dysponujemy fundamentem pod silną sztuczną inteligencję. Sytuację tę trafnie diagnozuje Erick J. Larson w *Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do*. Jego zdaniem z powodzeniem stosujemy aplikacje wyposażone w wąską inteligencję, które korzystają z mocy obliczeniowej cyfrowej technologii i dużej ilości danych, ale nie znaczy to, że dokonujemy stopniowego postępu, lecz raczej zbieramy nisko wiszące owoce. I na tym etapie prac nie ma przejścia od jednego do drugiego, nie istnieje żaden algorytm dla inteligencji uniwersalnej. Mamy także obiektywne powody, by sceptycznie podchodzić do tego, że taki algorytm powstanie w wyniku dalszych wysiłków nad systemami głębokiego uczenia⁴¹, chociaż wciąż jest nadzieja na wypracowanie takiego algorytmu⁴².

³⁸ *Sztuczna inteligencja non-fiction*, red. A. Obem, K. Szymielewicz, Fundacja Panoptykon, Warszawa 2020, s. 7.

³⁹ Por. J.R. S e a r l e, *Mind, Brains, and Program*, „The Behavioral and Brain Sciences” 3(1980) nr 3, s. 417; A.C. N e u b a u e r, *The Future of Intelligence Research in Age of Artificial Intelligence – With a Special Consideration of the Philosophical Movements of Trans- and Posthumanism*, „Intelligence” 87(2021), s. 5; L a r s o n, dz. cyt., s. 2.

⁴⁰ Zob. R. K u r z w e i l, *The Singularity is Near: When Humans Transcend Biology*, Viking Penguin, London, New York 2005.

⁴¹ Por. L a r s o n, dz. cyt., s. 2.

⁴² Zob. P. D o m i n g o s, *Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*, Basic Books, New York 2015.

Co można uznać za obiektywnie istniejącą trudność? Wydaje się, że problem tkwi w technicznym odtworzeniu jednego z kluczowych sposobów ludzkiego wnioskowania, które Charles S. Peirce określał mianem abdukcji⁴³. Ze względu na wagę tego zagadnienia w kontekście niniejszych rozważań jego opis zostanie nieco rozszerzony. Ten rodzaj rozumowania „polega na tym, by w obliczu zaskakującego zjawiska szukać hipotezy, z której można dedukcyjnie wyprowadzić to, że takie zjawisko zajdzie”⁴⁴. Jest to zatem testowanie hipotez poprzez próbkowanie możliwych przewidywań, które mogą być na nich oparte. Abdukcja jest potrzebna, gdy pojawiają się fakty sprzeczne z tym, czego powinniśmy się spodziewać. Wyjaśnienie musi być taką propozycją, która prowadziłaby do przewidywania zaobserwowanych faktów. Na przykład dzięki abdukcji możemy odpowiadać na pytania typu: czy woda przestała się lać, bo ktoś zakręcił kran, czy zbiornik się wyczerpał, a może chwilowo doszło do jakiegoś niekontrolowanego wycieku i woda cieknie, ale gdzie indziej? lub czy gdybym wtedy zachowała się inaczej, dzisiaj łatwiej byłoby mi zmierzyć się z tą sytuacją? Z technicznego punktu widzenia wnioskowanie próbujące przewidywać zdarzenia, w takim sensie „niezgodne z faktami”, bo ich zaistnienie jest jedynie rozważane, wymaga rozwijania równań algorytmicznych o dodatkowe składniki, które uwzględniałyby informacje o zmiennych pojawiających się w efekcie zmiany innych zmiennych. A to dotyczy przewidywania skutków działania i wyobraźni przeczącej faktom⁴⁵. Żadna maszyna obecnie nie dysponuje takim potencjałem⁴⁶.

Generalnie SI bazuje na stosowaniu znanych od starożytności metod pozyskiwania wiedzy, dedukcji (odgórna) oraz indukcji (oddolna). „Odgórne podejście odgrywało znaczącą rolę w pierwszych sztucznych inteligencjach, a w pierwszej dekadzie XXI wieku przeżyło nawrót w formie probabilistycznych (bayesowskich) modeli generatywnych. [...] Bayesowskie techniki ułatwiają określenie, która z dwóch hipotez jest bardziej prawdopodobna [...], ale żaden system nie potrafi efektywnie rozważyć wszystkich”⁴⁷. Oddolna metoda natomiast opiera się na technikach wykrywania statystycznych schematów w dużych zbiorach danych. Polega ona na zapewnieniu programowi dostępu do danych, z których wyłuskuje znaczące wzory. Rozwój SI bazujący na tej metodzie stał się możliwy

⁴³ Por. Ch.S. Peirce, *The Essential Peirce. Selected Philosophical Writings*, t. 2, (1993-1913), red. Nathan Houser, Indiana University Press, Bloomington, Indianapolis 1998, s. 95.

⁴⁴ A. Grobler, *Metodologia nauk*, Wydawnictwo Aureus, Wydawnictwo Znak, Kraków 2006, s. 102.

⁴⁵ Por. J. Pearl, *Ograniczenia nieprzejrzyście uczących się maszyn*, w: *Człowiek na rozdrożu. Sztuczna inteligencja – 25 punktów widzenia*, red. J. Brockman, tłum. M. Machnik, Helion, Gliwice 2020, s. 37n.

⁴⁶ Zob. Larson, dz. cyt.

⁴⁷ A. Gopnik, *AI kontra czterolatki*, w: *Człowiek na rozdrożu*, s. 246n.

dzięki technologicznemu dostępowi do danych, czyli zjawisku tak zwanego Big Data, a w tym przypadku kluczowe są korelacje⁴⁸.

Każde z tych dwóch podejść ma wady i zalety, które wzajemnie się uzupełniają. W metodzie oddolnej program nie wymaga wiedzy początkowej. Konieczny jest jedynie dostęp do danych. Z kolei podejście odgórne wymaga zaimplementowania dużej ilości „wiedzy początkowej”, która w przypadku każdego z programów okazywała się niewystarczająca. Stąd obecnie prace wielu badaczy z zakresu SI skupiają się na połączeniu obu metod⁴⁹. Jednak tym, co wydaje się najbardziej problematyczne, jest odkrywanie przez algorytmy głębokiego uczenia jedynie powierzchownych regularności statystycznych, a nie wysokopoziomowych pojęć abstrakcyjnych⁵⁰. Sztuczne sieci neuronowe nie osiągnęły jeszcze tego, co stanowi ważny aspekt ludzkiego uczenia się i rozumowania, a mianowicie możliwości tworzenia abstrakcyjnego modelu świata. Na razie SI jest wyposażona w filtry do rozpoznawania wzorów⁵¹ i ich specjalistycznego stosowania.

Ten, być może nieco przydługi, opis był konieczny, żeby uniknąć zjawisk, na które zwracają uwagę John Paul Mueller i Luca Massaron, czyli nierealistycznych wizji SI, rozbudzanych przez popkulturę, oraz ludzkiej skłonności do antropomorfizacji SI⁵², ujawniającej się szczególnie wtedy, kiedy tworzone rozwiązania techniczne z założenia mają odtwarzać ludzki sposób funkcjonowania. Dlatego trzeba powtórzyć, SI jest programem komputerowym opartym na szeregu algorytmów i póki co niczym więcej, nawet w sytuacji, gdy działania tego programu są nie do końca czytelne dla jego twórców, tak jak to się dzieje w przypadku głębokiego uczenia maszynowego⁵³.

⁴⁸ Zob. Ch. Anderson, *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, „Wired” 2008, <https://www.wired.com/2008/06/pb-theory/>; V. Mayer-Schönberger, K. Cukier, *Big data. Rewolucja, która zmieni nasze myślenie, pracę i życie*, tłum. M. Gładki, MT Biznes, Warszawa 2014; M. Kleppman, *Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable and Maintainable Systems*, O'Reilly, Beijing, Boston, Farnham, Sebastopol, Tokyo 2017.

⁴⁹ Por. Gopnik, dz. cyt., s. 247; H.L.J. van der Maas, L. Snoek, C.E. Stevenson, *How Much Intelligence Is There in Artificial Intelligence? A 2020 Update*, „Intelligence” 87(2021), s. 2-5.

⁵⁰ Por. J. Jo, Y. Bengi, *Measuring the Tendency of CNNs to Learn Surface Statistical Regularities*, Computer Science & Machine Learning, 30.09.2017, s. 8, <https://arxiv.org/abs/1711.11561>.

⁵¹ Por. S. Dehaene, *Jak się uczymy. Dlaczego mózgi uczą się lepiej niż komputery... jak dotąd?*, tłum. D. Rossowski, Copernicus Center Press, Kraków 2021, s. 63.

⁵² Por. J. Weizenbaum, *Computer Power and Human Reason. For Judgment to Calculation*, W.H. Freeman and Company, New York, San Francisco 1976, s. 6; J.P. Mueller, L. Massaron, *Sztuczna inteligencja dla bystrzaków*, tłum. K. Konatowicz, Helion, Gliwice 2021, s. 23.

⁵³ Por. Larson, dz. cyt., s. 23; Pearl, dz. cyt., s. 33-39; K.-F. Lee, *Inteligencja sztuczna, rewolucja prawdziwa. Chiny, USA i przyszłość świata*, tłum. K. Hejwowski, Media Rodzina, Poznań 2018, s. 20.

KOMUNIKACJA Z SI – WARUNKI BRZEGOWE

Wraz z konceptem myślącej maszyny, stopniowo rozwijanym do postaci, którą obecnie określamy mianem sztucznej inteligencji, pojawiały się pomysły na testowanie potencjału intelektualnego, jakim miałyby ona dysponować. Za pierwszy i najbardziej znany uznaje się test Alana Turinga. Wykorzystano w nim tak zwaną rozmowę z maszyną (ang. *imitation game*⁵⁴), w której człowiek nie orientuje się, że ma do czynienia z SI, a nie z drugim człowiekiem. Trudno się dziwić, że za najbardziej wiarygodny sprawdzian inteligencji przyjęto sytuację komunikacyjną, jak wskazywano bowiem wcześniej – komunikacja wymaga od człowieka przeprowadzenia skomplikowanych procesów percepcyjnych, interpretacyjnych, inteligibilnych, dokonywanych na bieżąco⁵⁵, pozwalających szybko dostosowywać się do wymogów sytuacji. Zdolność tę roboczo w niniejszym artykule określono mianem elastyczności komunikacyjnej. Człowiek osiągał ją na przestrzeni wieków, doskonaląc swoje umiejętności w tym względzie, ale także wypracowując rozwiązania techniczne, umożliwiające coraz bardziej zaawansowane jej formy. Mowa oczywiście o technologiach komunikacyjnych, takich jak: język, pismo, całe spektrum technologii elektrycznych, aż po technologię cyfrową⁵⁶.

Z punktu widzenia prowadzonych tu rozważań, zgodnie z przyjętymi wcześniej założeniami, kluczowe jest pytanie, jakie warunki komunikacyjne obecnie spełnia SI jako program komputerowy, wykorzystujący zaawansowane modele statystyczne do rozwiązywania zadań.

Odpowiedź na nie jest istotna z dwóch powodów. Po pierwsze, chodzi o uświadomienie sobie, jakim dysponujemy zakresem interfejsowym, pozwalającym odtwarzać (imitować) maszynie ludzką komunikację, żeby użytkownik nie musiał przyswajać sobie dodatkowych instrukcji, tylko wchodził w niejako „naturalne” dla siebie tryby komunikacyjne, które ułatwiają na przykład współpracę z maszyną⁵⁷. Po drugie, możemy wskazywać konkretne zastosowania dla

⁵⁴ Por. A.T. T u r i n g, *Computing Machinery and Intelligence*, „Mind” 49(1950) nr 236, s. 433- 460.

⁵⁵ R. D u n b a r, *Przyjaciele. O prawdziwej mocy naszych najważniejszych relacji*, tłum. T. Chawziuk, Copernicus Center Press, Kraków 2021, s. 128n.

⁵⁶ Por. G. O s i k a, *Tożsamości osobowa w epoce cyfrowych technologii komunikacyjnych*, Universitas, Kraków 2016, s. 35-99.

⁵⁷ Zob. S. H j o r t h, D. C h r y s o s t o m o u, *Human-robot Collaboration in Industry Environments: A Literature Review on Non-destructive Disassembly*, „Robotics and Computer-Integrated Manufacturing” 73(2022), s. 1-18; Y. L u, J.S. A d r a d o s, S.S. C h a n d, L. W a n g, *Humans Are Not Machines – Anthropocentric Human-machine Symbiosis for Ultra-flexible Smart Manufacturing*, „Engineering” 7(2021) nr 6, s. 734-737; Y. L u, H. Z h e n g, S. C h a n d, B. J i n s o n g, *Outlook on Human-centric Manufacturing Towards Industry 5.0*, „Journal of Manufacturing Systems” 62(2022), s. 612-627.

maszyn, bardziej świadomie przewidując konsekwencje ich użycia. Przykładowo już rozpoznano, że dzieci korzystające z Alexy⁵⁸ nabierają nawyków komunikacyjnych wymaganych do jej obsługi (chodzi o wydawanie komend, które następnie przenoszą na kontakty międzyludzkie⁵⁹). Z komunikacyjnego punktu widzenia taki efekt uboczny jest niepożądany, wprowadza bowiem do praktyk komunikacyjnych formy destrukcyjne. W tabeli 2 wyodrębnione wcześniej warunki zachodzenia komunikacji (por. tabela 1) opisano pod kątem zakresu możliwości, jakimi dysponuje SI. Roboczo sformułowano także ograniczenia powodujące brak spełnienia danego warunku, co oczywiście dotyczy aktualnej wiedzy.

Tabela 2. Obecny potencjał komunikacyjny SI

	Warunek zachodzenia komunikacji	Zdolność SI	Obecne ograniczenia SI
1.	Na proces komunikacyjny składa się: nadawca, przekaz, kanał, odbiorca oraz cel, dla którego się komunikujemy.	SI może inicjować komunikację z człowiekiem. Dowodem na to są funkcjonujące chat boty, począwszy od Elizy i A.L.I.C.E, po te, które otrzymały również wysokie noty w konkursie Hugh Loebnera ⁶⁰ , takie, jak: Rose, Mitsuku, Cleverbot, Elbot, Eugene Goostman, JFred, Ultra Hal ⁶¹ , Meena, DialoGPT, Xiaolce ⁶² , ale także LaMDA (aplikacja do projektowania chat botów) ⁶³ lub Project Debater, zajmujący się projektowaniem programów do uczestniczenia w debatach ⁶⁴ .	Istotny jest fakt, że do tej pory nie została przyznana nagroda główna w konkursie Loebnera; żaden chatbot nie zaimitował konwersacji ludzkiej. Nie znaczy to jednak, że programy nie mogą być interlokutorami, a więc pełnić funkcji nadawczych. Chociaż emitują komunikaty głównie językowe, mogą realizować cele komunikacyjne, do jakich zostały zaprojektowane i w jakimś sensie mogą być odbiorcami, czyli spośród dostępnego zbioru możliwych odpowiedzi mogą wybrać najbardziej relewantną.

⁵⁸ Alexa to tak zwany wirtualny asystent. Może on przeprowadzać interakcje głosowe, odtwarzać muzykę, sporządzać listy rzeczy do zrobienia, współpracować z domowymi systemami smart, strumieniowo przysyłać podcasty, odtwarzać audiobooki oraz dostarczać zaprogramowanych przez użytkownika informacji, na przykład o pogodzie, natężeniu ruchu czy sporcie. Alexa opiera się między innymi na technologiach automatycznego rozpoznawania mowy, przetwarzania języka naturalnego oraz innych formach słabej sztucznej inteligencji, kompatybilnej z jej oprogramowaniem.

⁵⁹ Por. Przeglądlińska, Oksanowicz, dz. cyt., s. 206n.

⁶⁰ Nagroda Loebnera – w konkursie ocenia się stopień imitowania rozmowy ludzkiej przez maszynę; Home Page of The Loebner Prize in Artificial Intelligence „The First Turing Test”, <https://web.archive.org/web/20101230195120/http://www.loebner.net/Prizetf/loebner-prize.html>.

⁶¹ Zob. H. Shah, K. Warwick, J. Vallverdú, D. Wu, *Can Machines Talk? Comparison of Eliza with Modern Dialogue Systems*, „Computers in Human Behavior” 58(2016), s. 278-295.

⁶² Zob. D. Adwardana, T. Long, *Towards a Conversational Agent that Can Chat About... Anything*, 28.01.2020, <https://ai.googleblog.com/2020/01/towards-conversational-agent-that-can.html>.

⁶³ Zob. B. Curtis, J. Savulescu, *Is Google's LaMDA Conscious? A Philosopher's View*, 15.06.2022, <https://theconversation.com/is-googles-lamda-conscious-a-philosophers-view-184987>; E. Collins, Z. Ghahramani, *LaMDA: Our Breakthrough Conversation Technology*, 3.05.2021, <https://blog.google/technology/ai/lamda/>.

2.	Każdy „znak” ma częściowo niestabilne znaczenie.	Dostęp do dużej ilości konwersacji, na jakich jest trenowana SI, oraz technologie automatycznego rozpoznawania mowy i przetwarzania języka naturalnego umożliwiają jej rozpoznawanie bardzo zróżnicowanych form użycia danego znaku, zatem niestabilność w jakimś zakresie może być odtwarzana, ale głównie w odniesieniu do komunikacji językowej ⁶⁵ .	SI nie rozumie słów, tylko dopasowuje wzorce konwersacyjne, jakimi dysponuje, które mają dla niej postać liczbową. Samo znaczenie wypowiedzi jest jej nieznane. „Niezdolność do rozumienia czegoś tak podstawowego jak słowa oznacza, że w tłumaczeniach z języków obcych generowanych przez sztuczną inteligencję, zawsze będzie brakowało tego czegoś, co jest niezbędne, aby oddać uczucia i myśli ukryte między słowami, a także wiernie przełożyć myśl. Słowa wyrażają myśl, a sztuczna inteligencja nie wie, co to jest” ⁶⁶ , nie potrafi też swobodnie korelować znaczenia z całego spektrum nośników znaczenia, w tym pozajęzykowych, dotyczy to także tak zwanej emocjonalnej SI (ang. Emotion AI), która skupia się na uczeniu komputerów rozpoznawania ludzkich emocji, ich ilościowego opisu oraz reagowania na emocje ⁶⁷ .
3.	Konieczność znajomości kontekstu dla ustalenia znaczenia przekazu.	Podobnie jak wyżej, dzięki technologiom automatycznego rozpoznawania języka oraz przetwarzania języka naturalnego na podstawie dostępu do dużej ilości danych, wskazujących na możliwe konteksty użycia danych znaków, SI może wybrać ten najbardziej adekwatny, ale tylko w ramach dostępnego zbioru danych, SI przetwarza bowiem dane i generuje wyniki, ale nie dokonuje interpretacji ⁶⁸ .	Jak wyżej, dopasowanie kontekstu odbywa się na podstawie wzorów konwersacyjnych, którymi dysponuje program. Mogą one zostać mu zaimplementowane „odgórnie” lub może wypracować je sam, ale nie zmienia to faktu, że nadal praca nad znaczeniem z uwzględnieniem kontekstu ma znaczenie – kontekst pozostaje poza jego zasięgiem, między innymi ze względu na brak wiedzy ⁶⁹ koniecznej do ustalenia kontekstu oraz brak możliwości jej stałej aktualizacji ⁷⁰ . Dlatego w aplikacjach SI pojawiają się pomysły wykorzystywania wiedzy już zgromadzonej na przykład w Wikipedii.

⁶⁵ Zob. I. Lauriola, A. Lavelli, F. Aiolfi, *An Introduction to Deep Learning in Natural Language Precessing: Models, Techniques, and Tools*, „Neurocomputing” 470(2022), s. 443-456.

⁶⁴ Zob. *Project Debater*, <https://research.ibm.com/interactive/project-debater/>.

⁶⁶ Mueller, Massaron, dz. cyt., s. 323; zob. Ö. Kocak, S. Park, Ph. Puranam, *Ambiguity Can Compensate for Semantic Differences in Human-AI Communication*, „Computers and Human Behavior” 6(2022), s. 1-21.

⁶⁷ Por. R. elKaliouby, C. Colman, *Girl decoded. Jak emocjonalna sztuczna inteligencja pomaga zrozumieć siebie i innych*, tłum. A. Jarosz, CeDeWu, Warszawa 2021, s. 17.

⁶⁸ Por. Mueller, Massaron, dz. cyt., s. 322.

⁶⁹ Por. Larson, dz. cyt., s. 53n.

⁷⁰ Por. tenże, dz. cyt., s. 202; Dehaene, dz. cyt., s. 65n.

4.	Miedzy uczestnikami komunikacji dokonuje się sprzężenie zwrotne pozwalające weryfikować przepływ informacji.	Sprzężenie zwrotne jest oczywiste z samego faktu wchodzenia w interakcję z maszyną, która dokonuje się w ramach typowego interfejsu podczas użytkowania.	Ograniczenia SI w tym względzie dotyczą interfejsu, jakim dysponuje maszyna, czy jest to na przykład obsługa głosowa czy taktynne panele sterowania.
5.	Ustalenie „czynności”, jakie wykonujemy, komunikując się.	W tym przypadku chodzi o zastosowanie maszyny zgodne z jej funkcjonalnością; wszystkie wąskie SI z zasady spełniają zakres czynności, do jakiego zostały zaprojektowane. Komunikacja dokonująca się między człowiekiem a SI ma charakter zadaniowy i jest pochodną jej użyteczności ⁷¹ .	W odniesieniu do tego warunku trudno mówić o ograniczeniach komunikacyjnych, jeżeli dochodzi do problemów, raczej są to wady projektowe, wpływające na funkcjonalność programu, oczywiście mowa o wąskiej SI.
6.	Warunek „wypowiedzenia” nie jest wystarczający do zachodzenia komunikacji.	Jak wskazywano wcześniej, SI wykorzystuje głównie, i to w ograniczonym zakresie, komunikację językową na poziomie generowania wypowiedzi. SI może przetwarzać numeryczne dane wejściowe za pomocą zbioru algorytmów i generować dane wyjściowe, nie będąc świadoma i nie rozumiejąc tego, co robi ⁷² .	Jak dotąd nie udało się przełamać trudności sformułowanej przez Johna R. Searle’a, opisaną jako argument chińskiego pokoju ⁷³ , który w uproszczeniu zakłada, że z faktu wymiany znaków (aktu lokucyjnego) nie możemy wnioskować o zachodzeniu czynności komunikacyjnej (aktu illokucyjnego). To, że chatbot mówi, nie znaczy, że wie, co robi, mówiąc, w przeciwieństwie do człowieka, który wypowiadając słowa „zachęcam cię”, faktycznie wykonuje czynność zachęcania.
7.	Realizowany cel komunikacyjny.	Ze względu na brak istnienia inteligencji intrapersonalnej w przypadku SI cel komunikacyjny jest zdeteminowany intencjami człowieka, dlatego w projektowaniu SI tak dużą rolę przywiązuje się do imitowania cech komunikacji ludzkiej aplikacji ⁷⁴ .	„Autorefleksja umożliwiającą zrozumienie własnych motywacji, a następnie wyznaczenie w oparciu o nie celów to rodzaj inteligencji, którą może się wykazać obecnie tylko człowiek” ⁷⁵ .
8.	Konieczność uwzględnienia interpretacji odbiorcy jako aktywnego uczestnika komunikacji.	Rozpoznana przez Grice’a tendencja do dokonywania wnioskowań asemantycznych powoduje, że w ludzkiej komunikacji, znając literalne znaczenie oraz dysponując wiedzą dotyczącą kontekstu, jeste-	W tym przypadku człowiek niezależnie od wkładu SI, będzie miał tendencję do „nadinterpretacji” potencjału komunikacyjnego sztucznej inteligencji. Być może właśnie taka sytuacja była przyczyną ogłoszenia faktu uzyskania świadomości

⁷¹ Por. J. van Dijk, *Spoleczne aspekty nowych mediów. Analiza społeczeństwa sieci*, tłum. J. Konieczny, Wydawnictwo Naukowe PWN, Warszawa 2010, s. 304n.

⁷² Por. Mueller, Massaron, dz. cyt., s. 28.

⁷³ Por. Searle, dz. cyt., s. 417-457.

⁷⁴ Zob. H. Jiang, Y. Cheng, J. Yang, S. Gao, *AI-powered Chatbot Communication with Customer: Dialogic Interactions, Satisfaction, Engagement, and Customer Behavior*, „Computers in Human Behavior” 134(2022), s. 1-14; K. Jinjun, K. Merrill, Ch. Collins, *AI as a Friend or Assistant: The Mediated Role of Perceived Usefulness in Social AI vs. Functional AI*, „Telematics and Informatics” 64(2021), s. 1-8.

⁷⁵ Mueller, Massaron, dz. cyt., s. 28.

		śmy w stanie nadpisać brakujące znaczenie, uzupełniając komunikacyjne braki za SI. Najlepszym tego przykładem jest przypadek opisywany z rozczarowaniem przez J. Weizenbauma, kiedy jego sekretarka angażowała się w rozmowy z ELIZĄ, czerpiąc z nich pełną satysfakcję komunikacyjną ⁷⁶ .	przez aplikację LaMDA, o którym informował jeden z inżynierów firmy Google. Swoją diagnozę oparł on na prowadzonych z programem rozmowach, które następnie opublikował ⁷⁷ . W przypadku SI natomiast odwzajemnienie wnioskowania asemantycznego wydaje się obecnie niemożliwe ze względu na wskazywany wcześniej brak umiejętności testowania hipotez poprzez próbkiwanie możliwych przewidywań, a właśnie tego wymagają wnioskowania asemantyczne.
9.	Uwspólnione działanie na poziomie percepcji, interpretacji, rozumienia.	W przypadku SI trudno jest mówić o konwergencji procesów percepcji, interpretacji i rozumienia w pełnym tych słów znaczeniu. Dostęp do danych i algorytmy, w tym głębokiego uczenia, nie są wystarczające do wykonania tak skomplikowanej pracy intelektualnej. Chociaż zaczynają być wypracowywane rozwiązania techniczne wykorzystujące niejednoznaczność języka do ustalania wspólnych znaczeń w trakcie konwersacji z człowiekiem, niejako włączania jego idiolektu do swoich zasobów ⁷⁸ .	Kluczowym ograniczeniem jest tutaj brak świadomości SI, a bez przełamania tej trudności ten warunek będzie mógł być realizowany tylko częściowo.
10.	Utrata konwencjonalnego nośnika znaczenia nie oznacza zaprzestania komunikacji; zaczynamy inne sygnały traktować jako nośniki, na przykład przy rozmowie przez telefon bardziej znaczące stają się walory wokalizacyjne.	SI pracuje na zaprogramowanych danych, dlatego przejście z jednego typu danych do innego musi być wcześniej zaprojektowane. Człowiek potrafi elastycznie czerpać z tych źródeł informacje, które w danym momencie wydają się najbardziej adekwatne, a następnie integrować nową wiedzę z tą, którą już dysponuje, i dokonywać re-kombinacji w celu rozwiązania problemu, w tym komunikacyjnego ⁷⁹ .	Na tym etapie rozwoju sieci neuronowych SI nie dysponuje takimi możliwościami ⁸⁰ .

⁷⁶ Por. W e i z e n b a u m, dz. cyt., s. 6.

⁷⁷ Zob. N. T i k u, *The Google Engineer Who Thinks the Company's AI Has Come to Life*, „Washington Post”, 11.06.2022, <https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/>; J. K w a ś n i e w s k i, *Studia nad historią i teorią cywilizacji*, <https://jacek.kwasniewski.org.pl>.

⁷⁸ Zob. K o c a k, P a r k, P u r a n a m, dz. cyt.

⁷⁹ Por. D e h a e n e, dz. cyt., s. 62-70.

⁸⁰ Por. tamże.

11.	Przebieg komunikacji wyzwalający z ograniczeń „tu” i „teraz” wymaga użycia medium, które traktujemy jak aktora społecznego.	W tym przypadku tendencja, która zostaje wyzwolona w trakcie kontaktów zapośredniczonych, związana z traktowaniem zarówno treści, jak i samego medium jako aktora społecznego, działa na korzyść SI, pozwala bowiem niwelować podziały na ludzki i nieludzki, mogące uruchamiać kategorie obcości ⁸¹ – toksyczne z punktu widzenia komunikacji.	Paradoksalnie traktowanie agentów nieludzkich jako aktorów społecznych może powodować brak realistycznej oceny, z czym człowiek ma do czynienia podczas interakcji z SI. Tak jak miało to miejsce w przytaczanych wcześniej przykładach ⁸² , kiedy de facto nie tyle dochodziło do skomunikowania, co w ludzkim mniemaniu taka komunikacja miała miejsce, jednak obiektywnie przekonanie to generowało dalsze poziomy nieporozumienia. Wydaje się, że na tym etapie prac nad SI nie mamy podstaw do myślenia, że maszyny mają coś więcej do powiedzenia, poza tym, co im podsuwamy ⁸³ .
12.	Komunikacja wymaga podzielenia uwagi i intencji.	Ten warunek jest spełniony tylko częściowo. SI potrafi w dużym stopniu „podać”, a być może raczej symulować podążanie uwagi za tym, na co swoją uwagę kieruje człowiek. Intencjonalność, jak wskazywano wcześniej, zależy od funkcjonalności programu, jaką nadaje mu projektant, ale także użytkownik, zatem można powiedzieć, że ludzki wkład w tym względzie jest decydujący.	Kategoria uwagi i intencjonalności stanowi istotną trudność projektową SI, pomimo że próbuje się te zagadnienia odłączać od kwestii związanych ze świadomością. Po pierwsze, wskazuje się na różne poziomy uwagi, od tej, która wymaga zaangażowania i ciągłego wysiłku oraz wspiera procesy uczenia i nabywania nowych umiejętności, przez mimowolną (ang. generally automatic) i szczególnie istotną dla zaangażowania emocjonalnego. Tego rodzaju uwagi nie można sprowadzić do procedur wyszukiwania i procesów selekcji, jakimi dysponuje SI, bo ona wymaga poziomów integracji poznawczej, które mogą okazać się zbyt trudne dla sztucznej inteligencji ⁸⁴ .

Źródło: Opracowanie własne.

⁸¹ Zob. Reeves, Nass, dz. cyt.

⁸² Zob. Weizenbaum, dz. cyt.; Tiku, dz. cyt.; Kwaśniewski, dz. cyt.

⁸³ Por. M. du Sautoy, *Kod kreatywności. Sztuka i innowacje w epoce sztucznej inteligencji*, tłum. T. Chawziuk, Copernicus Center Press, Kraków 2020, s. 327.

Przeprowadzone powyżej analizy pomagają nam uzmysłowić sobie zakresy komunikacyjne dostępne dla projektowanej obecnie SI i co ważniejsze – uświadomić sobie, jakich warunków brzegowych SI nie spełnia. Diagnoza ta może być pomocna w realnej ocenie potencjalnych zastosowań sztucznej inteligencji oraz edukacyjnych wyzwań, które przed nami stoją w kontekście powszechnego wykorzystania technologii kognitywnych w naszym życiu. Często wydaje się nam, że przełamywanie wykluczenia cyfrowego opiera się na przyswajaniu obsługi interfejsu, ale kompetencje digitalne to także świadomość ograniczeń technologii, którą się posługujemy⁸⁵. Ma to szczególnie duże znaczenie, kiedy znamy ludzkie tendencje nadinterpretacyjne ujawniające się w procesie komunikacji. Wydaje się, że jednym z podstawowych problemów, z którym mamy do czynienia w ludzkich kontaktach ze sztuczną inteligencją, jest nadmierna jej antropomorfizacja. SI przyjmuje dane na wejściu, analizuje je, odnajduje wzorce i generuje żądane dane wyjściowe, natomiast niczego nie rozumie. Jeżeli coś tworzy, to tylko w bardzo wąskim zakresie⁸⁶, nie ma wiedzy intrapersonalnej, zatem nie może wczuć się w niczyją sytuację⁸⁷, chociaż może u nas wywoływać takie wrażenie. Jest aplikacją zaprojektowaną przez programistę zgodnie z zakładaną funkcjonalnością. W przypadku komunikacji dobrze sprawdzają się powszechnie stosowane chatboty⁸⁸ – do wąskich zastosowań – na przykład w centrach serwisowych, serwisach zakupowych⁸⁹, do inicjowania kontaktu i wstępnej diagnozy problemów, trenowania konkretnej umiejętności komunikacyjnej, generowania rutynowych komunikatów, między innymi w poczekalniach, czyli tam, gdzie komunikacja ma charakter zadaniowy, z niskim poziomem skomplikowania. Jeżeli sytuacja komunikacyjna wymaga natomiast wykorzystania umiejętności z zakresu inteligencji emocjonalnej lub społecznej, efektywność SI z założenia jest bardzo ograniczona lub nieosiągalna. W ostatecznym rozrachunku SI ma duży potencjał

⁸⁴ Por. H.H. H a l a d j i a n, C. M o n t e m a y o r, *Artificial Consciousness and the Consciousness-attention Dissociation*, „Consciousness and Cognition” 45(2016), s. 214; D. G o l e m a n, *Focus. Sztuka koncentracji jako ukryte dążenie do doskonałości*, tłum. P. Szymczak, Media Rodzina, Poznań 2014, s. 8-28.

⁸⁵ Zob. G. O s i k a, *Dilemmas of Social Live Algorithmization: Technological Proof of Equity*, „Scientific Papers of Silesian University of Technology, Organization and Management Series” 2021, nr 151, s. 525-538.

⁸⁶ Por. S a u t o y, dz. cyt., s. 325-334.

⁸⁷ Por. M u e l l e r, M a s s a r o n, dz. cyt., s. 321.

⁸⁸ Por. K. W i e c z o r e k, *The Conscience of the Machin? Artificial Intelligence and the Problem of Moral Responsibility*, „Er(r)go. Teoria, Literatura, Kultura” 42(2021) nr 1, s. 20.

⁸⁹ Por. Y. Z h u, J. Z h a n g, J. W u, Y. L i u, *AI Is Better When I'm Sure: The Influence of Certainty of Needs on Consumer's Acceptance of AI Chatbots*, „Journal of Business Research” 150(2022), s. 642-652.

minimalizowania zadań, które będziemy musieli wykonywać, lecz nie będzie w żaden sposób przypominała człowieka⁹⁰.

Tym, co obecnie wydaje się kluczowe, jest zmiana sposobu komunikacji programistów na temat SI na szerszym forum społecznym, pozwalającym lepiej rozumieć, co kryje się za bardzo technicznym opisem technologii kognitywnych, nieczytelnym, a przez to w powszechnym odbiorze często mitologizowanym⁹¹.

Jak wskazano na początku artykułu, wobec strategii rozwojowych, jakie obecnie formułujemy, jednym z oczywistych założeń jest intensywna technicyzacja naszego życia oraz powszechne zastosowanie technologii kognitywnych. Wizje te wymuszają na nas konieczność rozwijania umiejętności, które pomogą nam optymalnie dostosowywać się do tych warunków. Jednym z aspektów tej sytuacji jest włączanie do życia społecznego aktorów nieludzkich, czyli tak zwanej sztucznej inteligencji. To znaczy, że musimy nauczyć się wchodzić w nowy typ interakcji i wypracowywać nowe rodzaje relacji. Niniejsze rozważania były próbą przybliżenia jednego z obszarów koniecznych do przemyślenia, kluczowego z punktu widzenia interakcji i relacji, jakim jest komunikacja. Założono, że najistotniejsze jest przebadanie warunków kluczowych dla zachodzenia komunikacji człowieka z SI. W związku z tym zaproponowano następujące pytania badawcze: jakie warunki brzegowe musi spełnić komunikacja oraz jakie warunki komunikacyjne obecnie spełnia AI jako program komputerowy wykorzystujący zaawansowane modele statystyczne do rozwiązywania zadań.

Na podstawie przeprowadzonych analiz ustalono warunki modelowe dla zachodzenia procesu komunikacji (por. tabela 1). Doprecyzowano pojęcie SI oraz rozpoznano, na jaki poziom „skomunikowania” człowieka z maszyną możemy obecnie liczyć (por. tabela 2). W badaniach wykorzystano analizę konceptualną, co pozwoliło rozwinąć teorię konieczną do operacjonalizacji potencjalnych prac empirycznych. W dalszej perspektywie chodziło także o bardzo szkicowe zarysowanie problemu związanego z ontologicznym wymiarem maszyn jako bytów, z którymi budujemy relacje, żeby pod tym względem móc dokonać bardziej świadomych osądów. W odniesieniu do tego przypadku optymalne wydawałoby się podejście postfenomenalistyczne, w którym relacja jest efektem mediacji dokonującej się w trakcie interakcji z maszyną i realistycznej oceny, czym ona de facto jest, oraz jakim zakresem jej funkcjonalności jesteśmy zainteresowani. To jednak wymaga zrozumienia samej technologii cyfrowej, dlatego wbrew oczekiwaniom, póki co w naszych postawach silniej zaznacza się tendencja instrumentalistyczna – przejawiająca się zbyt

⁹⁰ Por. D. S u m p t e r, *Osaczeni przez liczby. O algorytmach, które kontrolują nasze życie. Od Facebooka i Googla po fake newsy i bańki filtracyjne*, tłum. R. Kosarzycki, Copernicus Center Press, Kraków 2019, s. 301.

⁹¹ Zob. R. D a z e l e y i n., *Levels of Explainable Artificial Intelligence for Human-aligned Conversational Explanation*, „Artificial Intelligence” 299(2021), s. 1-29.

bezrefleksyjnym traktowaniem techniki lub substatywistyczna – skutkująca niebezpieczną antropomorfizacją SI.

BIBLIOGRAFIA / BIBLIOGRAPHY

- Adiwardana, Daniel, and Thang Luong. *Towards a Conversational Agent that Can Chat About... Anything*. January 28, 2020. <https://ai.googleblog.com/2020/01/towards-conversational-agent-that-can.html>.
- Anderson, Chris. *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, "Wired" 2008. <https://www.wired.com/2008/06/pb-theory/>.
- Arsovski, Slavko. *Quality of Life and Society 5.0*. "International Quality Conference 13 IQC Quality Research 2019." http://www.cqm.rs/2019/papers_iqc/81.pdf.
- Austin, John L. *How to do Things with Words*. London: Oxford University Press, 1962.
- Bar-Hillel, Yehoshua. "Indexical Expressions." *Mind, New Series*, no. 251 (1954): 359–79.
- Borgmann, Albert. *Technology and the Character of Contemporary Life: A Philosophical Inquiry*. Chicago: University of Chicago Press, 1984.
- Breque, Maija, Lars De Nul, Athanasios Petridis, and European Commission, Directorate-General for Research and Innovation. *Industry 5.0: Towards a Sustainable, Human-centric and Resilient European Industry*. Publications Office, 2021. <https://data.europa.eu/doi/10.2777/308407>.
- Brożek, Bartosz. *Granice interpretacji*. Kraków: Copernicus Center Press, 2014.
- Brynjolfsson, Eric, and Andrew McAfee. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. New York, London: W. W. Norton & Company, 2014.
- Collins, Eli, Zoubin Ghahramani. *LaMDA: Our Breakthrough Conversation Technology*. May 18, 2021. <https://blog.google/technology/ai/lamda/>.
- Craig, Robert T. "Communication Theory as a Field." *Communication Theory*, no. 9 (1999): 119–61.
- Curtis, Benjamin, and Julian Savulescu. *Is Google's LaMDA Conscious? A Philosophers' View*. June 15, 2022. <https://theconversation.com/is-googles-lamda-conscious-a-philosophers-view-184987>.
- Dance, Frank. *Human Communication Theory: Original Essays*. New York: Holt, Rinehart & Winston, 1967.
- Davies, Paul. *Demon w maszynie. Jak ukryte sieci informacji wyjaśniają tajemnicę życia*. Translated by Tomasz Lanczewski. Kraków: Copernicus Center Press, 2020.
- Dazeley, Richard, Peter Vamlew, Cameron Foale, Charlotte Young, Sunil Aryal, and Francisco Cruz. "Levels of Explainable Artificial Intelligence for Human-aligned Conversational Explanation." *Artificial Intelligence*, no. 299(2021): 1–29.

- Deguchi, Atsushi, Hirai Chiaki, Matsuoka Hideyuki, Tani Shigeyuki, Nakano Taku, and Oshima Kohel. "What is Society 5.0?" In Hitachi-UTokyo Laboratory, *Society 5.0: A People-centric Super-smart Society*. Singapore: Springer Open, 2020. https://doi.org/10.1007/978-981-15-2989-4_1.
- Dehaene, Stanislas. *Jak się uczymy: Dlaczego mózgi uczą się lepiej niż komputery... jak dotąd?*. Translated by Dariusz Rossowski. Kraków: Copernicus Center Press, 2021.
- van Dijk, Jan. *Społeczne aspekty nowych mediów. Analiza społeczeństwa sieci*. Translated by Jacek Konieczny. Warszawa: Wydawnictwo Naukowe PWN, 2010.
- Domingos, Pedro. *Master Algorithm: How the Quest for the Ultimate Learning Machine will Remake Our World*. New York: Basic Books, 2015.
- Dunbar, Robin. *Przyjaciele. O prawdziwej mocy naszych najważniejszych relacji*. Translated by Tadeusz Chawziuk. Kraków: Copernicus Center Press, 2021.
- Fielding, Guy, and Peter Hartley. "The Telephone: Neglected Medium." In *Studies of Communication*. Edited by Asher Cashdan, Martin Jordin. Oxford: Basil Blackwell, 1987.
- Fleischer, Michael. "Zarys ogólnej teorii komunikacji." In *Radykalny konstruktywizm. Antologia*. Translated by Bogdan Balicki, Dominik Lewiński, Bartosz Ryż, and Emil Szczerbuk. Wrocław: Centrum im. Willy Brandta, 2010.
- Freenberg, Andrew. *Transforming Technology: A Critical Theory Revisited*. New York: Oxford University Press, 2002.
- . "Critical Theory of Technology: An Overview." *Tailoring Biotechnologies*, no. 1(2005): 47–64.
- Gladden, Matthew. "Who will Be the Members of Society 5.0? Towards an Anthropology of Technologically Posthumanized Future Societies." *Social Science*, no. 148 (2019) 8(5): 1–39.
- Goleman, Daniel. *Focus. Sztuka koncentracji jako ukryte dążenie do doskonałości*. Translated by Piotr Szymczak. Poznań: Media Rodzina, 2014.
- Gopnik, Alison. "AI kontra czterolatek." In *Człowiek na rozdrożu. Sztuczna inteligencja – 25 punktów widzenia*. Edited by John Brockman. Translated by Marcin Machnik. Gliwice: Helion, 2020.
- Grice, Paul H. "Logic and Conversation." In *Studies in the Way of Words*. Cambridge, Massachusetts, London: Harvard University Press, 1967.
- Grobler, Adam. *Metodologia nauk*. Kraków: Wydawnictwo Aureus, Wydawnictwo Znak, 2006.
- Haladjian, Harry H., and Carlos Montemayor. "Artificial Consciousness and the Consciousness-attention Dissociation." *Consciousness and Cognition*, 45 (2016): 210–25.
- Hjorth, Sebastian, and Dimitrios Chrysostomou. "Human-robot Collaboration in Industry Environments: A Literature Review on Non-destructive Disassembly." *Robotics and Computer-Integrated Manufacturing*, vol. 73 (2022): 1–18.
- Hoffman, Christian H. "Is AI Intelligent? An Assessment of Artificial Intelligence, 70 Years after Turing." *Technology in Society*, 68 (2022): 1–12.

- Jaakkola, Elina. "Designing Conceptual Articles: Four Approaches." *AMS Review*, no. 10 (2020): 18–26.
- Jacobson, Roman. *Essais de linguistique generale*. Paris: Éditions de Minuit, 1963.
- Jiang, Hua, Yang Cheng, Yang Jeongwon, and Gao Shanbing. "AI-powered Chatbot Communication with Customer: Dialogic Interactions, Satisfaction, Engagement, and Customer Behavior." *Computers in Human Behavior*, no. 134 (2022): 1–14.
- Jinjun, Kim, Kelly Merrill, and Chad Collins. "AI as a Friend or Assistant: The Mediated Role of Perceived Usefulness in Social AI vs. Functional AI." *Telematics and Informatics*, no. 64 (2021): 1–8.
- Jo, Jason, and Yoshua Bengio. "Measuring the Tendency of CNNs to Learn Surface Statistical Regularities." *Computer Science & Machine Learning*, 30.09.2017, 1–13. <https://arxiv.org/abs/1711.11561>
- el Kaliouby, Rama, and Carol Colman. *Girl decoded. Jak emocjonalna sztuczna inteligencja pomaga zrozumieć siebie i innych*. Translated by Agnieszka Jarosz. Warszawa: CeDeWu, 2021.
- Kleppman, Martin. *Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable and Maintainable Systems*. Beijing, Boston, Farnham, Sebastopol, Tokyo: O'Reilly, 2017.
- Kocak, Özgecan, Sanghyun Park, and Phanish Puranam. "Ambiguity Can Compensate for Semantic Differences in Human-AI Communication." *Computers and Human Behavior*, no. 6 (2022): 1–21.
- Kulczycki, Emanuel. *Teoretyzowanie komunikacji*. Poznań: Wydawnictwo Naukowe UAM, 2012.
- Kurzweil, Ray. *The Singularity is Near: When Humans Transcend Biology*. New York: Viking Penguin, 2005.
- Kwaśniewski, Jan. *Studia nad historią i teorią cywilizacji*. <https://jacek.kwasniewski.org.pl>.
- Larson, Eric L. *The Myth of Artificial Intelligence. Why Computers Can't Think the Way We Do*. Cambridge, Massachusetts, London: The Belknap Press of Harvard University Press, 2021.
- Lasswell, Harold D. "The Structure and Function of Communication in Society." In *The Communication of Ideas*. Edited by Lyman Bryson. New York: Harper and Row, 1948.
- Lauriola, Ivano, Alberto Lavelli, and Fabio Aiolli. "An Introduction to Deep Learning in Natural Language Processing: Models, Techniques, and Tools." *Neurocomputing*, no. 470 (2022): 443–456.
- Lee, Kai-Fu. *Inteligencja sztuczna, rewolucja prawdziwa. Chiny, USA i przyszłość świata*. Translated by Krzysztof Hejwowski. Poznań: Media Rodzina, 2018 [e-book].
- Lu, Yuqian, Hao Zheng, Saahil S. Chand, and Bao Jinsong. "Outlook on Human-centric Manufacturing Towards Industry 5.0." *Journal of Manufacturing Systems*, no. 62 (2022): 612–27.

- Lu, Yuqian, Juvenal S. Adrados, Saahil S. Chand, and Lihui Wang. "Humans Are Not Machines – Anthropocentric Human-machine Symbiosis for Ultra-flexible Smart Manufacturing." *Engineering*, no. 7 (2021): 734–37.
- Luhmann, Niklas. *Systemy społeczne. Zarys ogólnej teorii*. Translated by Michał Kaczmarczyk. Kraków: Nomos, 2012.
- van der Maas, Han L.J., Lukas Snoek, Claire E. Stevenson. "How Much Intelligence Is There in Artificial Intelligence? A 2020 Update." *Intelligence* 87 (2021): 1–9.
- Manovich, Lev. *Język nowych mediów*. Translated by Piotr Cypryański. Warszawa: Wydawnictwa Akademickie i Profesjonalne, 2006.
- Mayer-Schönberger, Victor, and Kenneth Cukier. *Big data. Rewolucja, która zmieni nasze myślenie, pracę i życie*. Translated by Michał Gładki. Warszawa: MT Biznes, 2014 [e-book].
- Miczka, Tadeusz. "Człowiek techniczny, «człowiek rozszerzony», «człowiek 3D». Próba definicji." *Filo-Sofja* 17, no. 39(1) (2017): 37–47.
- Miller, Katherine. *Communication Theories. Perspectives. Processes and Context*. Boston: McGraw Hill, 2005.
- Mueller, John P., and Luca Massaron. *Sztuczna inteligencja dla bystrzaków*. Translated by Krzysztof Konatowicz. Gliwice: Helion, 2021.
- Nagroda Loebnera. <https://web.archive.org/web/20101230195120/http://www.loebner.net/Prizetf/loebner-prize.html>.
- Neubauer, Aljoscha C. "The Future of Intelligence Research in Age of Artificial Intelligence – With a Special Consideration of the Philosophical Movements of Trans- and Posthumanism." *Intelligence*, no. 87 (2021): 1–8.
- Obem, Anna, and Katarzyna Szymielewicz, eds. *Sztuczna inteligencja non-fiction*. Warszawa: Fundacja Panoptykon, 2020.
- Osika, Grażyna. "Czekając na osobliwości – o modelach interpretacji techniki." *Filo-Sofja* 17, 39(1)(2017): 65–78.
- . "Dilemmas of Social Live Algorithmization: Technological Proof of Equity", *Scientific Papers of Silesian University of Technology, Organization and Management Series*, no. 151 (2021): 525–38.
- . *Procesy i akty komunikacyjne. Koncepcje klasyczne i współczesne*. Kraków: Universitas, 2011.
- . "The Communicative Acts as Action." *Lingua as Communitas*, no. 18 (2008): 35–44.
- . *Tożsamości osobowa w epoce cyfrowych technologii komunikacyjnych*. Kraków: Universitas, 2016.
- Pearl, Judea. *Ograniczenia nieprzejrzyście uczących się maszyn*. In *Człowiek na rozdrożu. Sztuczna inteligencja – 25 punktów widzenia*. Edited by John Brockman. Translated by Marcin Machnik. Gliwice: Helion, 2020.
- Peirce, Charles S. *The Essential Peirce. Selected Philosophical Writings*. Vol. 2. (1893-1913). Edited by Nathan Houser et al. Bloomington, Indianapolis: Indiana University Press, 1998.

- Pleszczyński, Jan. *Epistemologia komunikacji medialnej. Perspektywa ewolucyjna*. Lublin: Wydawnictwo UMCS, 2013.
- Project Debater*. <https://research.ibm.com/interactive/project-debater/>.
- Przegalińska, Agnieszka, and Dariusz Jemielniak. *Społeczeństwo współpracy*. Warszawa: Wydawnictwo Naukowe Scholar, 2020.
- Przegalińska, Agnieszka, and Paweł Oksanowicz. *Sztuczna inteligencja. Nieludzka, arcyłudzka*. Kraków: Wydawnictwo Znak, 2020.
- Reeves, Byron, and Clifford Nass. *The Media Equation. How People Treat Computers, Television, and New Media Like Real People and Place*. New York: Cambridge University Press, 1992.
- Rogers, Everett M., and Lawrence D. Kinkaid. *Communications Networks: Toward a New Paradigm for Research*. New York: Free Press, 1981.
- du Sautoy, Marcus. *Kod kreatywności. Sztuka i innowacje w epoce sztucznej inteligencji*. Translated by Tadeusz Chawziuk. Kraków: Copernicus Center Press, 2020.
- Searl, John R. "Mind, Brains, and Programs." *The Behavioral and Brain Sciences* 3, 3 (1980): 417–57.
- . *Speech acts. An Essay in the Philosophy of Language*. Cambridge, New York, Melbourne, Madrid: Cambridge University Press, 1967.
- Shah, Huma, Kevin Warwick, Jordi Vallverdu, and Defeng Wu. "Can Machines Talk? Comparison of Eliza with Modern Dialogue Systems." *Computers in Human Behavior*, no. 58 (2016): 278–95.
- Shannon, Claude E., and Warren Weaver. *The Mathematical Theory of Communication*. Urbana: The University of Illinois Press, 1964.
- Suchacka, Małgorzata, Rafał Muster, and Mariusz Wojewoda. "Human and Machine Creativity: Social and Ethical Aspects of the Development of Artificial Intelligence." *Creativity Studies*, no. 14(2) (2021): 430–43.
- Sumpter, David. *Osaczeni przez liczący: O algorytmach, które kontrolują nasze życie. Od Facebooka i Googla po fake newsy i bańki filtracyjne*. Translated by Radosław Kosarzycki. Kraków: Copernicus Center Press, 2019.
- Śledziwska, Katarzyna, and Renata Włoch. *Gospodarka cyfrowa: Jak nowe technologie zmieniają świat*. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego, 2020.
- The World Manufacturing Forum Report. Skills for the Future of Manufacturing*. <https://worldmanufacturing.org/wp-content/uploads/WorldManufacturingFoundation2019-Report.pdf>.
- Tiku, Nitasha. "The Google Engineer Who Thinks the Company's AI Has Come to Life." *Washington Post*. June 11, 2022. <https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/>.
- Tomasello, Michael. *A Natural History of Human Thinking*. Cambridge: MIT Press, 2008.
- Turing, Alan T. "Computing Machinery and Intelligence." *Mind* 49 (1950): 433–60.
- Vallor, Shannon, ed. *The Oxford Handbook of Philosophy of Technology*. New York: Oxford University Press, 2022.

- Waismann, Friedrich. "Verifiability." *Journal of Symbolic Logic*, no. 12(3) (1947): 101.
- Walsh, Toby. *To żyje! Sztuczna inteligencja. Od logicznego fortepianu do zabójczej broni*. Warszawa: Wydawnictwo Naukowe PWN, 2018.
- Wielki słownik języka polskiego, s.v. "Warunek brzegowy." https://wsjp.pl/haslo/do_druku/97871/warunek-brzegowy.
- Weizenbaum, Joseph. *Computer Power and Human Reason: For Judgment to Calculation*. New York, San Francisco: W.H. Freeman and Company, 1976.
- Wendland, Michał. *Konstruktywizm komunikacyjny*. Poznań: Wydawnictwo Naukowe UAM, 2012.
- Wieczorek Krzysztof. "The Conscience of the Machine? Artificial Intelligence and the Problem of Moral Responsibility." *Er(r)go. Teoria, literatura, kultura*, no. 42 (2021): 15–34.
- Słownik języka polskiego PWN, s.v. "Zaimek zwrotny." <https://sjp.pwn.pl/slowniki/zaimek%20zwrotny.html>.
- Zhu, Yimin, Jiemin Zhang, Jifei Wu, and Yingyue Liu. "AI Is Better When I'm Sure: The Influence of Certainty of Needs on Consumer's Acceptance of AI Chatbots." *Journal of Business Research*, no. 150 (2022): 642–52.

ABSTRAKT / ABSTRACT

Grażyna OSIKA – Komunikacja człowieka ze sztuczną inteligencją. Warunki brzegowe

DOI 10.12887/36-2023-3-143-09

Wizje przyszłości formułowane w planowanych strategiach rozwoju za jeden z kluczowych czynników uznają intensywną technicyzację życia codziennego oraz powszechne wykorzystanie technologii kognitywnych. Strategie te wymuszają konieczność rozwijania nowych kompetencji. Jedną z nich jest komunikacja z nieludzkimi aktorami życia społecznego, takimi jak sztuczna inteligencja (SI), czyli nabywania umiejętności wchodzenia w nowy typ interakcji i rozwijania nowych rodzajów relacji. W niniejszych rozważaniach przyjmuje się, że najistotniejsze z tego punktu widzenia jest zbadanie kluczowych warunków zaistnienia komunikacji człowiek–SI, a więc ustalenie odpowiedzi na pytania: jakie warunki brzegowe muszą spełniać poszczególne sytuacje komunikacyjne, aby można je było uznać za komunikację oraz jakie warunki spełnia obecnie SI jako program komputerowy wykorzystujący zaawansowane modele statystyczne do rozwiązywania zadań? W wyniku przeprowadzonej analizy ustalono modelowe warunki występowania procesu komunikacji, zdefiniowano pojęcie SI oraz podjęto próbę rozpoznania na jaki poziom „komunikacji” między człowiekiem a maszyną możemy obecnie liczyć. Zastosowana w badaniach analiza konceptualna pozwoliła na wypracowanie teorii niezbędnej do operacjonalizacji potencjalnych prac empirycznych.

Słowa kluczowe: komunikacja, sztuczna inteligencja (SI), kompetencje komunikacyjne, Społeczeństwo 5.0, technologie kognitywne

Kontakt: Katedra Stosowanych Nauk Społecznych, Wydział Organizacji i Zarządzania, Politechnika Śląska, ul. Akademicka 2A, 44-100 Gliwice

E-mail: grazyna.osika@polsl.pl

ORCID: 0000-0002-8729-1001

Grażyna OSIKA, The Communication of Human Beings with Artificial Intelligence: Boundary Conditions

DOI 10.12887/36-2023-3-143-09

Visions of the future formulated in planned development strategies consider the intensive technicization of everyday life and the widespread use of cognitive technologies as key factors. These strategies force us to develop new competencies, one of which, resulting from the integration of nonhuman actors, such as artificial intelligence (AI), into social life, is the acquisition of the ability to enter into a new type of interaction and develop new types of relationships, that is, to communicate with AI. The discussion assumes that, from this point of view, it is the most relevant to examine the key conditions for the occurrence of human-AI communication, that is, to determine what boundary conditions must each communication situation meet in order to be considered communication and what conditions are currently met by AI as a computer program using advanced statistical models to solve tasks. As a result of the analysis, model conditions for the occurrence of the communication process were established, the concept of AI was adjusted, and an attempt was made to recognize what level of “communication” between man and machine can be currently achieved. The conceptual analysis used in the described research allowed the development of the theory necessary to operationalize potential empirical study.

Keywords: communication, artificial intelligence (AI), communication competence, Society 5.0, cognitive technology

Contact: Katedra Stosowanych Nauk Społecznych, Wydział Organizacji i Zarządzania, Politechnika Śląska, ul. Akademicka 2A, 44-100 Gliwice, Poland

E-mail: grazyna.osika@polsl.pl

ORCID: 0000-0002-8729-1001

Łukasz SAROWSKI
Małgorzata GRUCHOŁA

ROBOTYKA SPOŁECZNA W PERSPEKTYWIE (NIE)ANTROPOCENTRYZMU

Technologia sztucznej inteligencji stanowi bazę dla rozwoju robotyki społecznej. Jest ona stosunkowo nową dziedziną, zajmującą się próbą uchwycenia dynamicznej natury interakcji społecznych, w których istotną rolę mają odgrywać roboty humanoidalne. Bazuje na dorobku wielu dyscyplin, między innymi psychologii, medycyny, socjologii, nauk o komunikacji społecznej i mediach, nauk o poznaniu oraz uczenia maszynowego. Wielość ujęć definicyjnych oraz chaos semantyczny generują trudności w precyzyjnym opisie urządzeń, które mogłyby zostać objęte terminami „robot” czy „robot humanoidalny”.

Coraz większy udział zrobotyzowanych technologii w życiu społecznym implikuje konieczność opracowywania reguł określających sposób ich wykorzystania. Umiejętności komunikacyjne, poznawcze, sprawczość i rosnąca autonomiczność sprawiają, że roboty humanoidalne stają się niejako „nowymi” aktorami społecznymi, przeznaczonymi do współpracy z człowiekiem na nieco innych zasadach aniżeli te wyznaczane przez tradycyjne postrzeganie maszyny (bytu nie-ludzkiego) jako narzędzia służącego do rozwiązywania określonych problemów. Nowa sytuacja, wynikająca z dynamicznie rozwijającego się obszaru badań, którym jest robotyka, skłania do zadania wielu pytań dotyczących sposobu i skali wykorzystania jej wytworów, konsekwencji natury etycznej i społecznej oraz wiążących się z nimi implikacji prawnych – zwłaszcza wówczas, kiedy weźmiemy pod uwagę coraz szerszy zakres obdarzania robotów autonomicznością w kontekście podejmowania decyzji i aktywności społecznych. Nie bez znaczenia pozostaje również zjawisko antropomorfizacji robotów przez ich użytkowników, co może nieść ze sobą zasadnicze dylematy na polu społecznego wykorzystania sztucznej inteligencji. Problem ten ożywa zwłaszcza w ramach robotyki społecznej i procesów „uspołeczniania” jej wytworów (na przykład robotów towarzyszących).

W pierwszym przypadku można obserwować ożywienie dyskusji dotyczących prób opracowywania reguł funkcjonowania sztucznej inteligencji, które zapewnią bezpieczeństwo zarówno jej twórcom, jak i użytkownikom. W drugim przypadku można wskazać na dyskusje odnoszące się do prób konstruowania robotów humanoidalnych, imitujących społeczne zachowania człowieka, mogących modelować jego nastawienie (również reakcje emocjonalne). W konsekwencji można zaryzykować stwierdzenie, że obecnie ma miejsce proces osławiania sztucznej inteligencji (ang. Artificial Intelligence – AI) w postaci tak

zwanych robotów towarzyszących (ale nie tylko), co może doprowadzać do pewnego rodzaju „uspołecznienia” robotów, przyjmującego formy ich socjalizacji, analogiczne do procesów tak dobrze znanych socjologii w odniesieniu do człowieka.

Rodzi się pytanie o kryteria wdrażania ucieleśnionej sztucznej inteligencji w społeczeństwie. Przyjmujemy hipotezę, że proces „oswajania” technologii AI w postaci robotów humanoidalnych implikuje konieczność ich „uspołecznienia” we wszystkich wymiarach antropocentryzmu. Na potrzeby niniejszego artykułu przyjmujemy ujęcie antropocentryzmu zaproponowane przez Ewę Bińczyk¹.

Celem artykułu jest wykazanie – poprzez analizę literatury przedmiotu oraz wytworów robotyki społecznej – że proces „oswajania” robotów humanoidalnych wymusza opracowanie kryteriów i zasad ich funkcjonowania w życiu społecznym w konfrontacji z założeniami antropocentryzmu. Stąd analizie zostaną poddane jego cztery wymiary (poznawczy, ontologiczny, metodologiczny oraz aksjologiczny) w kontekście konieczności doprecyzowania przepisów etycznych, społecznych i prawnych dotyczących wykorzystania robotów humanoidalnych w społeczeństwie. Antropocentryzm poznawczy determinuje pytania o przepisy w zakresie „uspołeczniania” możliwości poznawczych zrobotyzowanych technologii. Antropocentryzm ontologiczny implikuje pytania o podmiotowość bytu nie-ludzkiego (robota humanoidalnego) w hierarchii bytów. Pytania o sprawczość i autonomiczność robotów generowane są przez antropocentryzm metodologiczny. Ostatni wymiar antropocentryzmu implikuje pytanie o kryterium aksjologiczne wykorzystywania i osławiania ucieleśnionej technologii AI, czyli o etykę robotów.

TECHNOLOGIA SZTUCZNEJ INTELIGENCJI W ROBOTYCE SPOŁECZNEJ

Obecnie nie ma powszechnie akceptowanej definicji technologii AI; w rzeczywistości wiele powiązanych pojęć, takich jak: roboty, uczenie maszynowe i AI, nakłada się na siebie². W literaturze przedmiotu sztuczna inteligencja ujmowana jest na wiele sposobów. Oznacza ona bowiem zarówno dziedzinę wiedzy, jak i jej wytwory (na przykład boty, chatboty, roboty). Marvin Minsky definiuje ją jako naukę o maszynach, które mogą realizować zadania wyma-

¹ Zob. E. Bińczyk, *Epoka człowieka. Retoryka i marazm antropocenu*, Wydawnictwo PWN, Warszawa 2018.

² Por. P. Lambert, *Computer Generated Works and Copyright: Selfies, Traps, Robots, AI and Machine Learning*, „European Intellectual Property Review” 2017, nr 39, s. 13; zob. M. Gruchola, *Technologia sztucznej inteligencji w dziennikarstwie a perspektywa deantropocentryzmu dziennikarza*, „Roczniki Nauk Społecznych” 14(2022) nr 2(50), s. 59-82.

gające ludzkiej inteligencji. Edward Feigenbaum podkreślił rolę celowości, uznając, że zasadniczym zadaniem jest instruowanie komputera, co ma zrobić, a nie jak ma dane zadanie wykonać³. Ray Kurzweil określił sztuczną inteligencję jako obszar badań zajmujący się naśladowaniem ludzkiej inteligencji przez maszynę. Obejmuje on między innymi systemy z bazą wiedzy, systemy ekspertowe, automatyczną naukę, rozpoznawanie obrazów, przetwarzanie języka naturalnego czy robotykę. Co więcej, traktowany jest jako sztuka tworzenia maszyn mogących wykonywać funkcje, które wymagają od człowieka inteligentnego działania⁴. Peter Menzel i Faith D'Aluisio zwracają uwagę, że sztuczna inteligencja jest sposobem programowania, pozwalającym komputerom działać samoczynnie, w tym uczyć się, rozumować, adaptować do otoczenia, korygować i ulepszać swoje działanie⁵. Należy także dokonać rozróżnienia technologii AI w postaci oprogramowania komputerowego, wykorzystywanej między innymi w dziennikarstwie zrobotyzowanym (algorytmicznym, zautomatyzowanym)⁶, i ucieleśnionej sztucznej inteligencji w postaci robotów humanoidalnych spotykanych w społeczeństwie (w domach opieki społecznej, portach lotniczych, agencjach medialnych i innych instytucjach).

Wielość prób definicyjnego ujęcia technologii AI, charakteryzującej się szerokim spektrum zastosowania, uwidacznia potrzebę opisu i uregulowania tej kwestii od strony społecznej (moralnej i etycznej) oraz prawnej w odniesieniu także do jej wytworów. Komisja Europejska zaproponowała definicję, która mogłaby stanowić podstawę do dalszych prac w powyższym zakresie. Określa ona sztuczną inteligencję, „odnosząc się do systemów, które wykazują inteligentne zachowanie poprzez analizę otoczenia i podejmowanie działań – z pewnym stopniem autonomii – w celu osiągnięcia określonych celów”⁷.

Technologia sztucznej inteligencji stanowi bazę do rozwoju robotyki społecznej. Jest ona stosunkowo nową dziedziną, zajmującą się próbą uchwycenia dynamicznej natury interakcji społecznych, w których istotną rolę mają odgrywać roboty humanoidalne. Bazuje na dorobku wielu dyscyplin, między innymi psychologii, medycyny, socjologii, nauk o komunikacji społecznej i mediach,

³ Definicje zostały przytoczone za: K. F i c o ń, *Sztuczna inteligencja nie tylko dla humanistów*, Bel Studio, Warszawa 2013, s. 81.

⁴ Por. R. K u r z w e i l, *The Age of Spiritual Machine*, Books Penguin, New York 1991, cyt. za: F i c o ń, dz. cyt., s. 82.

⁵ Por. P. M e n z e l, F. D' A l u i s i o, *Robo Sapiens: Evolution of a New Species*, MIT Press, Cambridge 2001, cyt. za: F i c o ń, dz. cyt., s. 82.

⁶ Por. G r u c h o ła, dz. cyt., s. 60-62; N. T h u r m a n, *Computational Journalism*, w: *The Handbook of Journalism Studies. Second Edition*, red. K. Wahl-Jorgensen, T. Hanitzsch, Routledge, New York 2019; K. N. D ö r r, *Mapping the Field of Algorithmic Journalism*, „Digital Journalism” 4(2016) nr 6, s. 701n.

⁷ Cyt. za: M. M i e r n i c k i, I. N g [Y. H u a n g], *Artificial Intelligence and Moral Rights*, „AI & Society” 36(2021) nr 2, s. 319.

nauk o poznaniu oraz uczenia maszynowego⁸. Wielość ujęć definicyjnych oraz chaos semantyczny generują trudności w precyzyjnym opisie urządzeń, które mogłyby zostać objęte terminami „robot” czy „robot humanoidalny”.

Janina Loh (za Catrin Misselhorn⁹) określa robota jako maszynę elektromechaniczną, mającą pewien zakres niezależności „cielesnej” pozwalającej działać jej w sposób względnie autonomiczny. „Składa się ona z co najmniej jednego procesora, który posiada czujniki zbierające informacje dotyczące otoczenia, oraz efekторы przetwarzające procesy na sygnały mechaniczne. Zachowanie robota jest lub przynajmniej wydaje się być autonomiczne, mające wpływ na otoczenie”¹⁰. Przytoczona definicja ma konotacje ściśle techniczne, stanowiące podstawę do dalszych definicyjnych ujęć, w tym zwłaszcza odnoszących się do społecznych aspektów wykorzystania zrobotyzowanych technologii. Inną propozycję rozumienia robota zawiera definicja robota społecznego, czyli „urządzenia zaprojektowanego w celu wywoływania społecznych interakcji za pomocą antropomorficznych kształtów (i komend wydawanych przy użyciu naturalnych sposobów komunikacji, takich jak gesty oraz mowa) z użytkownikami, którzy mogą przejawiać tendencje postrzegania robotów jako aktorów społecznych”¹¹. Kate Darling z kolei określa roboty społeczne jako „wcielone aktanty, które są autonomiczne, komunikujące się i wchodzące w interakcje na poziomie emocjonalnym”¹². Ich nadrzędnym wyróżnikiem powinny być zdolności do postępowania zgodnie z przyjętymi zasadami zachowań społecznych.

Modelowanie interakcji człowiek–robot na podstawie powszechnie obowiązujących norm społecznych otwiera pytania dotyczące ludzkich skłonności antropomorfizacyjnych w kontekście akceptacji robota jako towarzysza społecznego. Dotyczy to zwłaszcza tych robotów, które wykazują podobieństwo, zarówno w wyglądzie fizycznym (roboty humanoidalne), jak i w działaniach społecznych z udziałem i w stosunku do człowieka. Jak podkreślają Takayuki Kanda i Hiroshi Ishiguro, antropomorfizacja robotów może odgrywać zasad-

⁸ Por. Y. G u a n g - Z h o n g, D. P a o l o, K. D a n i c a, *Social Robotics – Trust, Learning and Social Interaction*, „Science Robotics” 3(2018) nr 21, s. 1n., <https://www.science.org/doi/epdf/10.1126/scirobotics.aau8839>.

⁹ Zob. C. Misselhorn, *Robots as moral agents*. w: *Ethics in Science and Society: German and Japanese Views*, red. F. Rovekamp, F. Bosse, Iudicium Verlag, Monachium 2013.

¹⁰ J. L o h, *Responsibility and Robot Ethics: A Critical Overview*, „Philosophies” 4(2019) nr 58, s. 7.

¹¹ K. M. L e e, N. P a r k, H. S o n g, *Can a Robot be Perceived as a Developing Creature: Effect of a Robot's Long-Term Cognitive Developments on its Social Presence and People's Social Responses Toward it*, „Human Communication Research” 2005, nr 31, s. 539.

¹² K. D a r l i n g, *Extending Legal Protection to Social Robots: The Effects of Anthropomorphism, Empathy, and Violent Behavior Towards Robotic Object*, University of Miami, Coral Gables 2012, s. 2.

niczą rolę w przypadku ich akceptacji jako towarzyszy ludzi (aktorów społecznych)¹³. W konsekwencji pojawia się potrzeba określenia statusu prawnego, a przez to i społecznego robotów. Szeroko dyskutowana jest propozycja objęcia najbardziej zaawansowanych maszyn statusem „osoby elektronicznej”¹⁴, co pozwalałoby na wprowadzanie określonych regulacji prawnych dotyczących nie tylko samych robotów, ale także ich użytkowników, konstruktorów, a nawet całych społeczeństw¹⁵.

ANTROPOCENTRYCZNE KRYTERIA „OSWAJANIA” SZTUCZNEJ INTELIGENCJI

Antropocentryzm uznaje prymat człowieka i tego, co ludzkie, nad innymi aktorami (bytami) nie-ludzkimi, także technologicznymi (na przykład zrobotyzowanymi maszynami)¹⁶. Obejmuje następujące komponenty: (1) poznawczy, zakładający konieczność przyjmowania perspektywy ludzkiej w interpretacji każdej sytuacji; (2) ontologiczny, przyznający człowiekowi wyjątkową pozycję w hierarchii bytów; (3) metodologiczny, koncentrujący się na sprawczej roli człowieka, dopuszczający jednak ingerencję bytów nie-ludzkich; (4) aksjologiczny, zakładający nadrzędność i bezdyskusyjność wartości ludzkich¹⁷.

KRYTERIUM POZNAWCZE

Antropocentryzm poznawczy podkreśla konieczność interpretowania wszelkich zagadnień z ludzkiego punktu widzenia zarówno podczas procesu rozumowania, jak i dokonywania wyborów moralnych, niezależnie od jednostkowego charakteru danej sytuacji. Generuje więc pytania o możliwości poznawcze robotów humanoidalnych, umiejętność podejmowania decyzji

¹³ Zob. T. Kanda, H. Ishiguro, *Human-Robot Interaction in Social Robotics*, CRC Press Taylor and Francis Group, Boca Raton 2013.

¹⁴ Zob. M. Delvaux, *Draft Report with Recommendations to the Commission on Civil Law Rules on Robotics*, „Committee on Legal Affairs” 2016, https://www.europarl.europa.eu/doceo/document/JURI-PR-582443_EN.pdf?redirect.

¹⁵ Zob. J. Hughes, *Citizen Cyborg: Why Democratic Society Must Respond to the Redesigned Human of the Future*, Basic Books, New York 2004.

¹⁶ Na temat teorii aktora-sieci Bruno Latoura, Michela Callona i Johna Lawa zob. B. Latour, *Reassembling the Social: An Introduction to Actor–Network–Theory*, Oxford University Press, New York 2005.

¹⁷ Por. B. Ińczyk, dz. cyt., s. 5n.; M. Szpunar, *Kultura algorytmów*, Instytut Dziennikarstwa, Mediów i Komunikacji Społecznej Uniwersytetu Jagiellońskiego, Kraków 2019, s. 89n.; Gruchola, dz. cyt., s. 66-77.

i dokonywania wyborów oraz ich konsekwencje, w niniejszym artykule traktowane jako kryteria „oswajania” technologii AI.

Powyższe pytania na gruncie robotyki społecznej pojawiają się w kontekście inspiracji ludzkimi funkcjami poznawczymi, które próbuje się odtwarzać w robotach. Dotyczą one przede wszystkim integracji procesów sensomotorycznych, będących podstawą do rozumowania czy też uczenia się przez doświadczenie. Przyjęcie paradygmatu ucieleśnionego umysłu miało za zadanie przezwyciężenie impasu wynikającego z piętrzących się problemów implementacji logiki formalnej do badań nad procesami poznawczymi systemów sztucznej inteligencji, które nie przyczyniły się do znaczącego postępu w budowie inteligentnych maszyn. Robotom brakowało – i nadal brakuje – podstawowej zdroworozsądkowej wiedzy uniemożliwiającej wyciąganie podstawowych wniosków. Zdano sobie sprawę, że w codziennych praktycznych rozumowaniach niezbyt często mamy do czynienia z dowodami, które można spotkać na gruncie matematyki. Stąd też zwrot w kierunku oparcia procesów poznawczych na ucieleśnieniu sztucznej inteligencji zapoczątkował nowy etap w badaniach nad interakcją sztucznego systemu poznawczego ze środowiskiem, w którym on działa, w tym z człowiekiem (robotyka społeczna)¹⁸. Naturalnym sposobem okazało się sięganie do analogicznych rozwiązań spotykanych w świecie przyrody, jak również aplikowanie wyników badań z zakresu społecznych interakcji międzyludzkich w obszar konstruowania robotów społecznych, w tym zwłaszcza robotów humanoidalnych, które zaczęto wyposażać w liczne sensory i receptory¹⁹. Miało to ułatwić robotom zbieranie informacji dotyczących otoczenia ich działania, stanowiących reprezentację faktów, nad którą miałyby one zacząć „myśleć”, a następnie przekładać wyniki „przemyśleń” na działania i próby zwerbalizowania ich w języku naturalnym. Stąd też obok badań nad procesami sensomotorycznymi można wskazać na te dotyczące przetwarzania języka naturalnego. Robot przetwarza język nie tylko w kontekście rozumienia wypowiedzi i przypisania jej odpowiedniego znaczenia, co w rezultacie ma doprowadzić do podjęcia adekwatnej decyzji i działania (należy podkreślić, że i tutaj dużą trudność sprawiają wyrażenia okazjonalne, funkcjonujące w języku potocznym, łączenie komunikacji werbalnej z pozawerbalnymi sygnałami oraz wieloznaczności dotyczące nie tylko samych słów, ale i wyrażen złożonych, na przykład amfibolii, elips, ekwiwokacji). „Zdolności komunikacyjne” robota działającego w określonym środowisku opierają się w dużej mierze na możliwości wiązania etykiet werbalnych

¹⁸ Zob. A. K i s i e l e w i c z, *Sztuczna inteligencja i logika*, Wydawnictwo Naukowe PWN, Warszawa 2017.

¹⁹ Zob. F. D e l c o m y n, *Biologically Inspired Robots*, w: *Bioinspiration and Robotics Walking and Climbing Robots*, red. M.K. Habib, I-Tech, Vienna 2007.

z przedmiotami znajdującymi się w polu jego działania²⁰. Generuje to znaczne trudności w konstruowaniu maszyn radzących sobie w dynamicznie zmiennym środowisku zarówno fizycznym, jak i społecznym. Rozwiązaniem problemu może być uczenie maszynowe (w tym: uczenie nadzorowane, półnadzorowane, nienadzorowane, uczenie na bieżąco, uczenie ze wzmocnieniem)²¹.

Z pytaniem o możliwości poznawcze robotów wiąże się problem ich autonomiczności. Ogólnie możliwości te można określić jako warunki pozwalające na sprawne działanie w środowisku fizycznym, które umożliwiają rozwiązywanie problemów bez ingerencji człowieka. Takie rozumienie autonomiczności wiąże się jednak z pewnymi ograniczeniami²². O ile na gruncie rozważań antropocentrycznych za byt autonomiczny możemy przyjąć ten, który posiada możliwość rozwijania wolnej woli, co w konsekwencji czyni go niezależnym od czynników zewnętrznych w procesie podejmowania decyzji, o tyle w odniesieniu do autonomii bytów nie-ludzkich uwidacznia się pewne uproszczone jej rozumienie²³. Jest ono konsekwencją próby uniknięcia wklęcia się w dyskusje dotyczące posiadania przez roboty wyżej wspomnianej dyspozycji w takim stopniu, w jakim przysługuje ona człowiekowi²⁴. Widoczne są w tym przypadku nawiązania do etologii i rozumienia autonomiczności poprzez odniesienie jej do sprawności działania zwierzęcia w zmiennych warunkach otoczenia. Nie pomija się w tym przypadku zagadnienia świadomości i samoświadomości, choć ich znaczenie jest mniejsze niż w kontekście dyskusji antropocentrycznych.

W przypadku budowania zaawansowanych robotów zauważalne są inspiracje paradygmatem ucieleśnionego umysłu. Warto odnotować, że wiele w tym zakresie wniosły badania fenomenologiczne²⁵, zwłaszcza w kontekście zainspirowania się badaczy światem przyrody, skutkującego pojawieniem się tak zwanego podejścia Biologically Inspired Robotics²⁶, zakładającego ko-

²⁰ Zob. L. S t e e l s, *Social Learning and Verbal Communication with Humanoid Robots*, s. 1n., https://www.researchgate.net/publication/228800852_Social_learning_and_verbal_communication_with_humanoid_robots.

²¹ Por. R. S. M i c h a ł s k i, *O naturze uczenia się – problemy i kierunki badawcze*, „Informatyka” 1988, nr 2, s. 4-7.

²² Por. M. C z a r n o c k a, *Podmiot poznania a nauka*, Wydawnictwo Naukowe UMK, Toruń 2012, s. 249.

²³ Por. C. S c h m i d t, F. K r a e m e r, *Robots, Dennett and the Autonomous: a Terminological Investigation*, „Minds and Machines” 16(2006) nr 1, s. 73-80.

²⁴ Por. R. K o p e ć, *Autonomia systemów bojowych*, „Przegląd Geopolityczny” 2016, nr 17, s. 134n.

²⁵ Zob. A. P r z e g a l i ń s k a, *Istoty wirtualne. Jak fenomenologia zmieniała sztuczną inteligencję*, Wydawnictwo Universitas, Kraków 2016.

²⁶ Zob. R. D. B e e r, i in., *Biologically Inspired Approaches to Robotics*, „Communications of the ACM” 40(1997) nr 3, s. 31-38.

nieczność posiadania przez roboty ciał organizmopodobnych. Rodzaj ucieleśnienia robotów będzie determinował sposób ich działania i radzenia sobie z napotykanymi problemami na zasadzie analogii obserwowanych w świecie zwierząt. Podobny postulat można odnaleźć w krytyce odcieleśnionych sztucznych sieci neuronowych Huberta Dreyfusa na rzecz ich ucieleśnienia, na przykład w postaci humanoidalnego robota, jeżeli ten miałby przejawiać zachowania podobne do człowieka²⁷.

Należy zauważyć, że jedną z zasadniczych kwestii pojawiających się w powyższym kontekście jest pytanie o zakres autonomicznego działania takiego robota, co rodzi potrzebę dokonania odróżnienia działań autonomicznych od zautomatyzowanych w sztucznych systemach poznawczych, w których występują inspiracje światem przyrody. Zasadniczym kryterium tego rozróżnienia nie są wolność wyboru i samodzielność decyzji, ale rutynowość i czasowość działania. W przypadku urządzeń autonomicznych, którymi mają być roboty, podkreśla się konieczność analizowania przez system posiadanego poziomu energii potrzebnej do realizacji określonych działań bez konieczności odwoływania się do ludzkiej interwencji. System sam powinien „rozeznąć”, czy wykonanie zadania jest możliwe, jaki poniesie on koszt energetyczny i czy istnieją alternatywne sposoby realizacji zadania pozwalające zaoszczędzić zgromadzony zasób energii. Takie działanie charakterystyczne jest dla badań autonomii biologicznej, którą można określić jako pewnego rodzaju samozachowawczą cechę organizacyjną organizmu, umożliwiającą zarządzanie interakcjami ze światem w celu podtrzymania jego żywotności. Granica między działaniem zautomatyzowanym a autonomicznym opierać się będzie na zdolności do samokierowania i podejmowania decyzji, a więc na zdolnościach adaptacyjnych, które będą miały zasadnicze znaczenie w przypadku społecznego usytuowania robota i konieczności podejmowania przez niego decyzji²⁸.

Powyższe uwagi można odnieść do trzech sposobów ujmowania autonomii robotów. Pierwszy z nich postrzega ją jako krańcową automatyzację – maszyna może wykonywać większość operacji bez interwencji człowieka przy jednoczesnym podkreśleniu, że ma ona ograniczenia narzucone w procesie programowania. W drugim przypadku może stanowić pojęcie przeciwstawne do automatyzacji – instrukcje wprowadzone do programu stanowią element wyjściowy umożliwiający zarządzanie funkcjonowaniem maszyny (nie pozwalają one jednak człowiekowi na jej pełne kontrolowanie w procesach decyzyjnych, które są warunkowane przez zmienne bodźce środowiskowe). Trzecie ujęcie zwraca uwagę, że automatyzacja jest pierwszym elementem

²⁷ Por. T. Ziemięke, *Czym jest to, co zwiemy ucieleśnieniem*, „Avant” 6(2015) nr 3, s. 166n.

²⁸ Por. Kopeć, dz. cyt., s. 136.

autonomii – system poznawczy stopniowo wzbogacony zostaje o możliwość samodzielnego podejmowania decyzji²⁹.

Zagadnienie podejmowania decyzji przez sztuczne systemy poznawcze i wiążące się z tym odpowiedzialności stanowią obecnie jedną z najbardziej naglących kwestii do rozwiązania przez środowiska naukowe. Warto w tym miejscu zaprezentować skalę Thomasa B. Sheridana oraz Williama Verplanka³⁰, która została zaproponowana co prawda w drugiej połowie dwudziestego wieku, ale nadal ma duży potencjał eksplikacyjny, wyznaczający zakres uczestnictwa oprogramowania komputerowego, a obecnie technologii AI w podejmowaniu decyzji. Przedstawia się ona następująco: (1) człowiek wykonuje całą pracę, a następnie implementuje ją i jej wynik do komputera; (2) AI pomaga w zadaniu poprzez określenie opcji możliwości jego rozwiązania; (3) AI pomaga nie tylko w określeniu opcji możliwości rozwiązania problemu, ale również sugeruje jedną z nich; (4) AI wybiera opcję rozwiązania problemu, która następnie zostaje (nie zostaje) zrealizowana przez człowieka; (5) AI wybiera opcję rozwiązania problemu i realizuje ją po autoryzacji człowieka; (6) AI wybiera opcję rozwiązania problemu, realizuje ją i informuje człowieka w odpowiednim czasie, kiedy ma ją zatrzymać; (7) AI wybiera opcję rozwiązania problemu, realizuje ją i informuje człowieka, co zostało wykonane; (8) AI wybiera opcję rozwiązania problemu, realizuje ją i informuje człowieka, co zostało wykonane, jeśli człowiek o to poprosi; (9) AI wybiera opcję rozwiązania problemu, realizuje ją i informuje człowieka, co zostało wykonane, oraz decyduje, czy człowiek powinien być poinformowany; (10) AI wybiera opcję rozwiązania problemu, realizuje ją, jeśli zdecyduje, że operacja musi zostać przeprowadzona, i informuje człowieka, co zostało wykonane, jeśli zdecyduje, że człowiek powinien być o tym poinformowany³¹.

Powyższa skala pozwala dookreślić wcześniej zarysowane w kontekście badań etologicznych kryteria autonomiczności robotów: niezależność, asertywność i adaptacyjność. Ponadto pokazuje trudności we współpracy człowieka z programem komputerowym, co uniemożliwia jednoznaczne udzielenie odpowiedzi na pytanie, kto ponosi odpowiedzialność za podjęcie określonej decyzji, zwłaszcza jeżeli będzie ona wiązała się z poważnymi konsekwencjami.

²⁹ Por. D.G. Johnson, M.E. Norman, *Responsibility Practices in Robotic Warfare*, „Military Review”, May-June 2014, s. 12-21; cyt. za: Kopeć, dz. cyt., s. 136n.

³⁰ Zob. T.B. Sheridan, W.L. Verplank, *Human and Computer Control of Undersea Teleoperations: Technical Report on Contract N00014-77-0256*, https://www.researchgate.net/publication/23882567_Human_and_Computer_Control_of_Undersea_Teleoperatr.

³¹ Por. R. Parasuraman, T. Sheridan, Ch. Wickens, *A Model for Types and Levels of Human Interaction with Automation*, „IEEE Transactions on Systems, Man and Cybernetics – Part A: Systems and Humans” 36(2000) nr 3, s. 287; cyt. za: Kopeć, dz. cyt., s. 139n.

W tym kontekście nasuwają się kolejne pytania, czy rosnący poziom autonomicznego działania robotów może być przesłanką do ujęcia ich w ramy podmiotu moralnego i czy w ogóle można mówić o sytuacji moralnej w przypadku działania robota. John P. Sullins wyróżnia trzy zasadnicze stanowiska w tej kwestii. Pierwsze z nich traktuje sytuację moralną jako iluzję. Błędnie bowiem przypisujemy prawa oraz obowiązki maszynom, co może wynikać z ludzkich skłonności antropomorfizacyjnych, które mogą być wzmacniane rosnącym zakresem niezależnego działania robotów. Drugie stanowisko wskazuje, że sytuacja jest pseudomoralna, gdyż robotom nie można przypisywać własności, które pozwoliłyby traktować je jako podmioty moralne. Trzecie stanowisko jest najbardziej liberalne. Zakłada, że o ile w ramach robotyki mamy do czynienia z nowymi problemami, które wcześniej nie były poruszane w takiej skali jak obecnie, to postawienie robota w określonej moralnej sytuacji będzie w pełni moralne bez względu na okoliczności³².

Człowiek, jeszcze do niedawna, najczęściej postrzegał zaawansowane technologie jedynie jako narzędzia mediujące między ich użytkownikiem a celem działania, bez uwzględniania ich w ocenie moralnej. Sytuacja ta ulega jednak zmianie z racji wspomnianej już wielokrotnie rosnącej autonomiczności narzędzia, które „aspirują” do bycia postrzeganymi jako „współsprawcy” działania³³. Współsprawstwo z kolei nasuwa pytanie dotyczące odpowiedzialności dzielonej pomiędzy użytkownika i narzędzie. John P. Sullins w tym kontekście przytacza porównanie do sytuacji z psem przewodnikiem osoby niewidomej, zadając pytanie, komu powinna należeć się pochwała moralna za jego poprawne zachowanie. Odpowiedź na tak postawione pytanie wiąże ze sobą zarówno właściciela zwierzęcia, trenera, który wyszkolił psa, ale również i samego psa. Zwolennicy tego typu argumentacji zwracają uwagę na analogiczność, która może występować między konstruktorem, użytkownikiem i robotem. Oznacza to, że sytuacja moralna każdorazowo powinna być rozpatrywana z punktu widzenia złożoności interakcji między zaprogramowanym robotem, środowiskiem oraz użytkownikiem. Odpowiedzialność robota będzie warunkowana pełnioną przez niego rolą społeczną. To ona wyznaczy zakres uprawnień i obowiązków, które powinny być przez niego przestrzegane³⁴.

³² Por. J.P. Sullins, *When is a Robot a Moral Agent?*, „International Review of Information Ethics” 6(2006) nr 12, s. 24.

³³ Por. K. Krzysztofek, *Sprawczość ludzka, transludzka i postludzka w społeczeństwie nasyconym technologicznie*, w: *Moc sprawcza ludzi i organizacji*, red. L.W. Zacher, Wydawnictwo Poltext, Warszawa 2016, s. 330.

³⁴ Por. J. Marx, Ch. Tiefensee, *Of Animals, Robots and Men*, „Historical Social Research” 40(2015) nr 4, s. 85.

KRYTERIUM ONTOLOGICZNE

Antropocentryzm ontologiczny, przyznający człowiekowi uprzywilejowaną pozycję i rolę w hierarchii bytów, generuje pytania o pozycję i „podmiotowość” robotów humanoidalnych, czyli bytów nie-ludzkich, w hierarchii wszystkich bytów. Odrzuca on możliwość przypisania podmiotowości robotom społecznym, zakładając, że tylko człowiek jest podmiotem moralnym, istotą rozumną, refleksyjną i samoświadomą³⁵.

Roboty, będące bytami nie-ludzkimi, są powodem wielu problemów i pytań badawczych związanych z podmiotowością i „osobowością”. Obowiązujące dokumenty Unii Europejskiej nie regulują jeszcze tej kwestii, chociaż podejmowane są działania w kierunku wypracowania pojęcia tzw. osobowości elektronicznej i objęcia przynajmniej niektórych wytworów robotyki i sztucznej inteligencji podmiotowością prawną. W tym kontekście zasadne jest ich ujęcie w obszar analiz kontekstu społecznego. Randall Collins zawiązuje podmiotowość do sprawczości i interaktywności³⁶. W antropocentryzmie podmiot społeczny, będąc podmiotem komunikacji, określa narzędzia kształtowania i rozumienia wspólnej przestrzeni znaczeń (kodowania i dekodowania) oraz współtworzy przestrzeń społecznych relacji uczestników procesu komunikacji, włącznie z możliwością doświadczenia drugiego człowieka. Możliwość ta w robotach humanoidalnych ogranicza się tylko do interaktywności, definiowanej jako zdolność wzajemnego oddziaływania na siebie różnych systemów oraz układów urządzeń, mogących wymieniać się informacjami i modyfikować wzajemnie swoje działania³⁷. Warto zauważyć, że już takie zastosowanie ucieleśnionej technologii sztucznej inteligencji (jej antropomorfizowania za pomocą humanoidalnego ucieleśnienia umożliwiającego interakcje na poziomie werbalnym i pozawerbalnym) może zmienić również status ontologiczny przyznający człowiekowi uprzywilejowaną pozycję w hierarchii bytów. Według Magdaleny Szpunar³⁸ coraz bardziej zasadne jest pisanie o deantropocentryzacji humanistyki, którą należy rozumieć jako odejście od humanistycznej wizji człowieka, traktującej go jako miarę wszechrzeczy oraz centrum zainteresowań badawczych na rzecz zrobotyzowanych maszyn. Ich osławianie generuje (nie)antropocentryzm.

³⁵ Por. Gruchola, dz. cyt., s. 68-73.

³⁶ Por. R. Collins, *Interaction Ritual Chains*, Princeton University Press, Princeton 2004, s. 79.

³⁷ Por. M. Wierzbchoń, M. Łukowska, *Ucieleśnione poznanie*, w: *Przewodnik po kognitywistyce*, red. J. Bremer, Wydawnictwo WAM, Kraków 2016, s. 605-622.

³⁸ Por. Szpunar, dz. cyt., s. 31.

Oswajaniu zrobotyzowanych maszyn sprzyja także ucieleśnienie organizmoidalne, które determinuje sposób postrzegania robota³⁹. Zakłada ono umieszczenie AI w ciałach fizycznych, zwłaszcza organizmoidalnych, charakteryzujących się formą i zdolnościami sensomotorycznymi zbliżonymi do żywych organizmów ludzkich (między innymi w robotach humanoidalnych) bądź organizmów zwierzęcych (w animaloidach)⁴⁰. Komponent techniczny ucieleśnienia uwzględniający sensomotoryczność umożliwia fizyczne „zaistnienie” oraz funkcjonowanie robota w środowisku zewnętrznym, społecznym (między innymi w domach opieki społecznej, portach lotniczych) oraz sprawczość. Komponent społeczno-psychologiczny – zgodnie z koncepcją „doliny niesamowitości” (ang. Uncanny Valley) Masahiro Moriego⁴¹ – może stanowić barierę w budowaniu relacji człowiek–robot, a tym samym jej oswajaniu. Komponent antropologiczny zaś łączy dwa przywołane powyżej wymiary w celu zbudowania autonomicznych robotów społecznych⁴² (na przykład roboty humanoidalne Hiroshiego Ishiguro, roboty: Sophia, Cog, Pepper).

KRYTERIUM METODOLOGICZNE

Antropocentryzm metodologiczny ogniskuje się na roli i sprawczości człowieka, dopuszczając ingerencję bytów nie-ludzkich, a więc i robotów. Generuje pytania o sprawczość robotów humanoidalnych (uznanie ich autonomiczności zakłada przypisanie im pewnej formy sprawczości i intencjonalności) oraz faktyczny zakres autonomiczności ich działania⁴³. Na wstępie należy podkreślić różnicę między tradycyjnym pojmowaniem autonomiczności w perspektywie personalistycznej⁴⁴, nawiązującej do zagadnienia wolnej woli i dyspozycji podmiotu do jej rozwijania, a zagadnieniem stopniowania autonomiczności robotów humanoidalnych w procesach decyzyjnych przeprowadzanych na podstawie informacji odbieranych za pomocą systemów odbiorczych – autonomiczności mogącej sugerować intencjonalność działania⁴⁵. Autonomiczność

³⁹ Por. Ziemię, dz. cyt., s. 162n.

⁴⁰ Por. tamże, s. 167; Ł. S a r o w s k i, *Robot humanoidalny jako podmiot czy przedmiot interakcji społecznej?*, XI Polski Zjazd Filozoficzny, Lublin, 9-14.09.2019 r.

⁴¹ Por. M. M o r i, *The Uncanny Valley*, „Energy” 7(1970) nr 4, s. 33-35.

⁴² Zob. M. K l i c h o w s k i, *Narodziny cyborgizacji. Nowa eugenika, transhumanizm i zmierzchność edukacji*, Wydawnictwo Naukowe UAM, Poznań 2014.

⁴³ Por. G r u c h o ł a, dz. cyt., s. 73-75.

⁴⁴ Por. S. K a m i Ń s k i, B. C z u p r y n, *Hasło Autonomia*, w: *Powszechna encyklopedia filozofii*, t. 1, red. A. Maryniarczyk, Polskie Towarzystwo Tomasza z Akwinu, Lublin 2000, s. 421.

⁴⁵ Por. D e l v a u x, dz. cyt., s. 1.

warunkuje więc pewne formy sprawczości, a tym samym i pozory intencjonalności w działaniach społecznych.

W antropocentryzmie pojęcie sprawczości odnoszone jest do sfery intencjonalnych (noszących znamiona emocjonalności oraz racjonalności) działań człowieka. Przez wiele lat czynnik sprawczy kojarzony był jedynie z człowiekiem jako jednostką obdarzoną rozumem, świadomością oraz możliwością działania. Wyraźne było odróżnienie podmiotu działania od przedmiotu, na który działanie było kierowane. W miarę upowszechniania się technologii sztucznej inteligencji, zwłaszcza kiedy zyskała ona pewien rodzaj autonomiczności i sprawstwa, pojęcie to zostało rozmyte⁴⁶.

KRYTERIUM AKSJOLOGICZNE

Jeśli antropocentryzm aksjologiczny traktuje wartości, priorytety oraz cele, ku którym zmierza człowiek, jako nadrzędne i niepodlegające dyskusji, oznacza on podporządkowywanie otaczającej rzeczywistości interesom ludzkim⁴⁷. Generuje więc pytanie o kryterium aksjologiczne wykorzystania i osławiania technologii AI, czyli o etykę robotów, która jest nową formą etyki stosowanej, subkategorią etyki maszyn. Gianmarco Veruggio definiuje ją następująco: „[Jest to] etyka stosowana, której cel stanowi opracowanie naukowych/kulturowych/technicznych narzędzi, które mogą być wykorzystywane przez różne grupy społeczne. Narzędzia te mają na celu promowanie i zachęcanie do rozwoju robotyki dla dobra społeczeństwa i jednostek, a także zapobieganie jej niewłaściwego wykorzystania przeciwko ludzkości”⁴⁸.

Pierwsze symposium poświęcone tej problematyce odbyło się w styczniu 2004 roku we Włoszech⁴⁹. Wezwano wówczas środowiska akademickie (filozofów, socjologów, psychologów, antropologów, etyków, prawników i innych naukowców zajmujących się konstruowaniem robotów) do udziału w tworzeniu podstaw etyki ujmującej w swoje ramy dylematy wynikające z konstruowania i upowszechniania coraz bardziej zrobotyzowanych technologii. Jeszcze w tym samym roku podczas Międzynarodowych Targów Robotów w Fukuoce w Japonii ich uczestnicy podpisali *Światową deklarację robotów*

⁴⁶ Por. Krzysztofek, dz. cyt., s. 331.

⁴⁷ Por. Bińczyk, dz. cyt., s. 5.

⁴⁸ G. Veruggio, *The Birth of Roboethics*, w: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2005)*, Workshop on Robot Ethics, Barcelona, 2005, s. 1-4, <https://philpapers.org/rec/VERTBO-3>. Tłum. – Ł.S., M.G.

⁴⁹ Zob. First International Symposium on Roboethics: The Ethics, Social, Humanitarian and Ecological Aspects of Robotics, Villa Nobel, Sanremo, Italy, 30-31 January 2004, <http://www roboethics.org/sanremo2004/>.

[„World Robot Declaration”], w której zobowiązali się do produkcji robotów „przyjaznych” dla człowieka, a w szczególności robotów: (1) będących partnerami współistniejącymi z ludźmi; (2) pomagających ludziom zarówno w zakresie realizacji zadań fizycznych, jak i psychicznych; (3) przyczyniających się do budowania bezpiecznego i pokojowego społeczeństwa⁵⁰. Niniejsze trzy wskazania odnoszą się do wielu obszarów wykorzystania zrobotyzowanych technologii, między innymi w gospodarce, gdzie toczy się debata nad pozytywnymi i negatywnymi skutkami zastępowania w pracy człowieka robotami; nad efektami społecznych interakcji ludzi z robotami, mogącymi wywoływać w nich pozytywne i negatywne emocje; nad opieką zdrowotną, która miałaby zostać oparta na robotach; nad dostępem do danych zbieranych przez roboty i nad możliwościami ich niewłaściwego wykorzystania (na przykład w celach terrorystycznych). Badania nad etyką robotów znajdują więc swoje uzasadnienie na gruncie nauk prawnych, w których rośnie potrzeba opracowania odpowiednich regulacji. Są one niezbędne zarówno w przypadku dostarczenia wytycznych co do wykorzystania zrobotyzowanych technologii w wyżej przytoczonych obszarach, jak i uregulowania jednej z kluczowych kwestii, którą jest odpowiedzialność wynikająca z działania autonomicznego robota. Problem ten generuje wiele dylematów wiążących się z ustaleniem podmiotu (konstruktor, użytkownik, robot) odpowiedzialnego za efekty powyższych działań.

Obok dyskusji popierających potrzebę ustanowienia etyki robotów pojawiają się głosy sprzeciwu. Zdaniem Janiny Loh odnoszą się one do dwóch zarzutów. Pierwszy dotyczy braku konkretnego przedmiotu analizy, etyka bowiem nie zajmuje się przedmiotami nieożywionymi. Drugi wskazuje, że włączenie sztucznych systemów poznawczych w obszar analiz nie spowoduje konieczności ich szczególnego traktowania. Obydwie te kwestie stanowią analogię zagadnień podejmowanych w ramach tradycyjnych ujęć etycznych. W odpowiedzi na powyższe zarzuty Loh podkreśla, że mogą one dotyczyć każdej etyki ograniczonej do konkretnego kontekstu (na przykład etyki środowiskowej, etyki zwierząt). Ponadto etyka robotów formułuje uzasadnione pytania dotyczące przesłanek moralnego sprawstwa, wartości moralnych, w które powinno się wyposażać systemy sztucznej inteligencji, jeśli te w przyszłości miałyby współpracować z człowiekiem na zasadach „partnerstwa”. Pozwala także badaczom na nowo odkrywać podstawowe pytania dotyczące rozumienia natury zła i dobra, implikujące różnorodne relacje człowieka z zaawansowanymi technologiami. Jednocześnie należy zauważyć, że konieczność podejmowania takiej refleksji widoczna jest już od dłuższego czasu w projektach literackich, między innymi w zbiorach opowiadań Isaaca Asimova (w tym

⁵⁰ Cyt. za: Veruggio, dz. cyt., s. 3.

*Ja-robot*⁵¹, w którym zawarł trzy prawa robotyki⁵²) czy w powieści Philipa K. Dicka *Blade Runner. Czy androidy śnią o elektrycznych owcach?*⁵³. Także Stanisław Lem⁵⁴ wiele uwagi poświęca uwikłaniu człowieka w skomplikowane relacje ze swoimi wytworami. Elementem łączącym niemal wszystkie opracowania literackie jest rosnące podobieństwo robotów do człowieka, poszerzający się zakres ich autonomicznego działania, implikujący ich niezależność, która może pozostawać poza kontrolą człowieka.

Problemy generowane przez robotykę i sztuczną inteligencję wskazują na trzy zasadnicze wymiary dyskusji etycznej. Pierwszy odnosi się do samych robotów i pytania o skonstruowanie wytycznych dla ich działania. Drugi dotyczy opracowywania reguł i wskazań dla konstruktorów robotów, a trzeci obejmuje samych użytkowników.

Pierwszy wymiar ogniskuje problemy związane z opracowywaniem konkretnych reguł etycznych działania autonomicznych robotów. Przykładem jest wielokrotnie przytaczany w literaturze przedmiotu z zakresu etyki sztucznej inteligencji „dylemat wagonika” opisany przez Philippe Foot⁵⁵. Potencjalne – zebrane przez Thomasa Cathcarta – odpowiedzi na pytanie, jak powinien zachować się motorniczy z punktu widzenia różnych teorii etycznych, stanowią inspirację dla możliwych rozwiązań analogicznych problemów etycznych zrobotyzowanych technologii. „Programiści – jak słusznie zauważa Grzegorz Szulczewski – stają zatem przed realnym problemem (na kogo skierować nadjeżdżający pojazd) i oczekują od etyków tego, co wydaje się niemożliwe, a mianowicie jednoznacznej, niepozostawiającej wątpliwości odpowiedzi. Dzieje się tak również dlatego, że wybór konkretnej teorii etycznej służącej do analizy problemu moralnego określanego jako dylemat wagonika wyznacza szereg możliwych rozwiązań”⁵⁶. Stąd obecnie jesteśmy w stanie udzielać raczej ogólnych wskazań, niż podawać konkretne rozwiązania, które mogłyby

⁵¹ Zob. I. A s i m o v, *Ja-robot*, tłum. Z.A. Królicki, Wydawnictwo Rebis, Poznań 2015.

⁵² Tamże, s. 46. Prawo pierwsze głosi, że robot nie może wyrządzić krzywdy człowiekowi ani przez brak reakcji dopuścić, aby człowiekowi stała się krzywda; prawo drugie – robot musi wykonać każdy rozkaz człowieka z wyjątkiem rozkazów kolidujących z pierwszym prawem; prawo trzecie – robot musi chronić swoją egzystencję, o ile nie jest to sprzeczne z pierwszym i drugim prawem.

⁵³ Zob. P.K. D i c k, *Blade Runner. Czy androidy śnią o elektrycznych owcach?*, tłum. S. Kędzierski, Rebis, Poznań 2016.

⁵⁴ Zob. S. L e m, *Bajki robotów*, Biblioteka Gazety Wyborczej, Warszawa 2008.

⁵⁵ Zob. T. C a t h c a r t, *Dylemat wagonika*, tłum. K. Bażyńska-Chojnacka, Dom Wydawniczy PWN, Warszawa 2014. Wymykający się spod kontroli motorniczego wagon kolejowy zbliża się do rozwidlenia bez możliwości zatrzymania. Motorniczy może skierować wagon na boczny tor lub pozostać na linii głównej. Sytuację komplikuje fakt, że na linii głównej torów znajduje się pięć osób, na torze bocznym jedna osoba. Motorniczy musi więc zdecydować, czy poświęci życie pięciu osób, czy jednej.

⁵⁶ G. S z u l c z e w s k i, *Sztuczna inteligencja a inteligencja moralna. Zagadnienia wstępne cybernetyki*, „Annales: Ethics in Economic Life” 22(2019) nr 3, s. 22.

zostać zaimplementowane w robotach społecznych, co jednak nie rozwiązuje problemu. Konieczne wydaje się opracowywanie regulacji dotyczących obecności coraz bardziej zrobotyzowanych maszyn w życiu człowieka oraz reguł, którymi powinny się one kierować w działaniach niezależnych od człowieka. Nie mniej istotną kwestię stanowią rekomendacje dla producentów robotów. Można przywołać kodeks EURON (European Robotics Research Network), który zaleca zobligowanie konstruktorów do: (1) zabezpieczenia kontroli robotów przez ludzi, a więc ograniczania ich pełnej autonomiczności (Safety: Ensure Human Control of Robot); (2) konieczności wyeliminowania sytuacji, w których robot mógłby być wykorzystywany w celach nielegalnej działalności (Security: Prevent Wrong or Illegal Use); (3) ochrony oraz przechowywania danych przez roboty (Privacy: Protect Data Held by Robot); (4) monitorowania działania robotów (Traceability: Record Robot's Activity); (5) nadawania unikalnych identyfikatorów (Give Unique ID to each Robot)⁵⁷.

Trzeci obszar dylematów natury etycznej dotyczy samych użytkowników w sytuacjach, w których zaawansowanie technologiczne maszyn może być wykorzystywane do działań niezgodnych z przyjętymi regułami postępowania czy regulacjami prawnymi. Z jednej strony technologia AI i robotyka mogą być narzędziami cyberprzestępstwa (na przykład trollingu prawnoautorskiego⁵⁸), z drugiej jednak sam robot może być przedmiotem niewłaściwego użycia i zachowania ze strony użytkownika. Warto odnotować, że problem ten zauważony został już w piśmiennictwie prawniczym⁵⁹. Wpisuje się on w szerszy kontekst badań nad robotami społecznymi, zwłaszcza humanoidalnymi – podejmuje się próby opracowywania ich charakterystyki „płciowej”⁶⁰ oraz podnosi się problem intymnych relacji ludzi z robotami, co rodzi wiele pytań i kontrowersji⁶¹.

*

Celem artykułu było wykazanie – poprzez analizę literatury przedmiotu oraz wytworów robotyki społecznej – że proces „oswajania” robotów humanoidalnych wymusza opracowanie kryteriów i zasad ich funkcjonowania

⁵⁷ Por. R. C a m p a, *Kodeksy etyczne robotów: zagadnienie kontroli sprawowanej przez człowieka*, „Pomiary Automatyka Robotyka” 2011, nr 3, s. 89.

⁵⁸ Zob. N. M i l e s z y k, *Copyright trolling – skala zjawiska i propozycje rozwiązań*, Centrum Cyfrowe, <https://centrumcyfrowe.pl/czytelnia/copyright-trolling-skala-zjawiska-i-propozycje-rozwiazan/>.

⁵⁹ Por. M. P a w i ł s k a, *Dopuszczalność gwałtu na humanoidalnym robocie*, „Acta Universitatis Lodzensis. Folia Iuridica” 2019, nr 86, s. 9-18.

⁶⁰ Por. R. S ø r a a, *Mechanical Genders: How do Humans Gender Robots*, „Gender, Technology and Development” 21(2017) nr 1-2, s. 99-115.

⁶¹ Zob. K. D e v l i n, *Seksroboty. O porządku, nauce i sztucznej inteligencji*, Wydawnictwo Dolnośląskie, Wrocław 2018.

w życiu społecznym w konfrontacji z założeniami antropocentryzmu. Wnioski z analizy czterech wymiarów antropocentryzmu w ramach robotyki społecznej pozwalają na częściowe potwierdzenie przyjętej we wprowadzeniu hipotezy badawczej, zakładającej, że proces osławajania technologii AI w postaci robotów humanoidalnych implikuje konieczność ich „usławocznienia” we wszystkich wymiarach antropocentryzmu wyróżnionych przez Ewę Bińczyk.

Analiza literatury przedmiotu prowadzi do następujących wniosków. Niezbędna jest konceptualizacja takich pojęć, jak: robot, robot społeczny, robot humanoidalny, robotyka społeczna z perspektywy antropocentrycznej (humanistycznej i społecznej). Ujęcia spotykane w literaturze przedmiotu najczęściej koncentrują się na wymiarze technicznym. Należy także doprecyzować przepisy i kryteria diagnostyczne regulujące status prawny wytworów robotyki społecznej (podmiot czy przedmiot). Przyznanie podmiotowości prawnej, która wiązałaby się również ze zmianą ich statusu społecznego, determinowałoby także próby zmiany, roli człowieka w społeczeństwie, a nawet zrównywania z nim robotów humanoidalnych.

Z analizy literatury przedmiotu wynika, że rozwój robotyki społecznej zmienia sposób postrzegania i analizowania zjawisk społecznych, w których coraz większą rolę odgrywają technologie zrobotyzowane. Realizują one zadania do niedawna zarezerwowane tylko dla człowieka, pełniąc różne funkcje społeczne (asystenci w domach opieki społecznej: robot *Paro*, wykładowcy, studenci, asystenci edukacyjni, interaktywni towarzysze dnia codziennego). I chociaż na płaszczyźnie ontologicznej nie zostały jeszcze wypracowane formalne kryteria warunkujące podmiotowość robotów i nie przyznano im podmiotowości prawnej, to intensywny rozwój technologii, zmieniając pozycję robotów w przestrzeni społecznej, sprzyja nieformalnemu, uznaniowemu przyznawaniu im podmiotowości. Należy zwrócić uwagę na świadomość ograniczonej możliwości realizowania rozmaitych zadań przez roboty, co nie zmienia jednak faktu, że technologie te mogą, a może już determinują społeczne zmiany dotyczące nie tylko sfery praktycznej ich wykorzystania w przemyśle, gospodarce, medycynie, ale również sfery relacji społecznych opartych na emocjonalnym zaangażowaniu poprzez konstruowanie robotów potrafiących odczytywać ludzkie emocje i adekwatnie na nie reagować (roboty *Pepper* oraz *Kismet*). Można zaryzykować stwierdzenie, że przyznanie statusu prawnego, a co się z tym wiąże uznania ich „poza przedmiotowej roli społecznej” będzie generować zmiany we wszystkich obszarach antropocentryzmu; tym samym będzie on warunkował proces „osławajania” i usławoczniania ucieleśnionej sztucznej inteligencji.

Ontologiczny (nie)antropocentryzm oznaczałby co najmniej zrównanie człowieka z robotem humanoidalnym. Chociaż w hierarchii bytów człowiek jako jednostka rozumna, podmiot moralny wciąż zajmuje wysoką pozycję, to

w Unii Europejskiej są już podejmowane działania legislacyjne nad przyznaniem robotom „osobowości elektronicznej”. Ucieleśnienie organizmoidalne sztucznej inteligencji (choć nie zawsze konieczne, na przykład w przypadku robotów niehumanoidalnych) umożliwia jej „zaistnienie” w świecie bytów materialnych, otwiera perspektywy wchodzenia w kontakt z człowiekiem, budowania relacji.

Sprawczość sztucznej inteligencji, będąca pochodną automatyzacji działania, przemawia za (nie)antropocentryzmem w sferze społecznej. Z przeprowadzonej analizy wynika, że dostrzegalna jest potrzeba poszerzenia kategorii bytu nie-ludzkiego, aktora, a nawet podmiotu życia społecznego o zrobotyzowane technologie, które mogą znacząco determinować sposób ludzkiego funkcjonowania. Przykładem takiego stanu rzeczy są pytania dotyczące konsekwencji osvajania ludzi z technologiami interaktywnymi, które w postaci robotów społecznych (humanoidalnych, animaloidalnych) będą obecne w różnych sferach ludzkiej aktywności społecznej. W tym zakresie najwięcej pytań generują jednak roboty humanoidalne z racji swego podobieństwa do człowieka. Próba odtwarzania społecznej natury ludzkiej w maszynie pociągać będzie za sobą wiele problemów, które wymagają podjęcia dyskusji nad ich przyszłymi konsekwencjami. Etyka robotów wpisuje się w tym wymiarze w obszar refleksji stanowiących podstawę do opracowania regulacji określających zakres stosowalności i odpowiedzialności omawianych w artykule robotów i ich społecznego statusu.

Należy zauważyć, że w ocenie procesu „osvajania” sztucznej inteligencji w postaci tak zwanych robotów humanoidalnych ani nadmierny entuzjazm, ani też skrajny sceptycyzm nie mają uzasadnienia. W niemal każdej sferze życia społecznego pozostaje on zarówno szansą, jak i zagrożeniem. Jak odnieść się zatem do tych przeciwstawnych poglądów? Bez wątpienia oba przyjęte stanowiska zawierają część prawdy, budząc i obawę, i nadzieję.

Po analizie literatury przedmiotu nasuwają się pewne postulaty i rekomendacje. Perspektywę badań w przyszłości wyznaczają: (1) problem deantropomorfizacji humanistyki i jednocześnie antropomorfizacji techniki; (2) zjawisko ucieleśnienia organizmoidalnego bytów nie-ludzkich oraz ucieleśnienia technicznego dla bytów ludzkich (między innymi technika przelewu umysłu – ang. mind uploading); (3) najnowsze formy komunikacji oraz interaktywności (między innymi Telepathic Web czy Symbionet Web).

BIBLIOGRAFIA / BIBLIOGRAPHY

- Asimov, Isaac. *Ja-robot*. Translated by Zbigniew A. Królicki. Poznań: Wydawnictwo Rebis, 2015.
- Beer, Randall D., et al. "Biologically Inspired Approaches to Robotics". *Communications of the ACM* 40, no. 3 (1997): 30–38.
- Bińczyk, Ewa. *Epoka człowieka. Retoryka i marazm antropogenu*. Warszawa: Wydawnictwo PWN, 2018.
- Campa, Ricardo. "Kodeksy etyczne robotów: zagadnienie kontroli sprawowanej przez człowieka." *Pomiary Automatyka Robotyka*, no. 3 (2011): 86–90.
- Cathcart, Thomas. *Dylemat wagonika*. Translated by Katarzyna Bażyńska-Chojnacka. Warszawa: Wydawnictwo PWN, 2014.
- Collins, Randall. *Interaction Ritual Chains*. Princeton: Princeton University Press, 2004.
- Czarnocka, Małgorzata. *Podmiot poznania a nauka*. Toruń: Wydawnictwo Naukowe UMK, 2012.
- Darling, Kate. *Extending Legal Protection to Social Robots: The Effects of Anthropomorphism, Empathy, and Violent Behavior Towards Robotic Object, We Robot Conference*. Coral Gables: University of Miami, 2012.
- Delcomyn, Fred. "Biologically Inspired Robots." In *Bioinspiration and Robotics Walking and Climbing Robots*. Edited by Maki K. Habib. Vienna: I-Tech, 2007.
- Delvaux, Mady. "Draft Report with Recommendations to the Commission on Civil Law Rules on Robotics". In *Committee on Legal Affairs* 2016, https://www.europarl.europa.eu/doceo/document/JURI-PR-582443_EN.pdf?redirect.
- Devlin, Kate. *Seksroboty. O pożądaniu, nauce i sztucznej inteligencji*. Wrocław: Wydawnictwo Dolnośląskie, 2018.
- Dick, Philip K. *Blade Runner. Czy androidy śnią o elektrycznych owcach?* Translated by Sławomir Kędzierski. Poznań: Wydawnictwo Rebis, 2016.
- Dörr, Konstantin Nicholas. "Mapping the Field of Algorithmic Journalism." *Digital Journalism* 4(2016) no. 6: 700–22.
- Ficoń, Krzysztof. *Sztuczna inteligencja nie tylko dla humanistów*. Warszawa: Bel Studio, 2013.
- Gruchola, Małgorzata. "Technologia sztucznej inteligencji w dziennikarstwie a perspektywa deantropocentryzmu dziennikarza." *Roczniki Nauk Społecznych* 14, no. 2(50) (2022): 59–82.
- Guang-Zhong, Yang, Paolo Dario, and Danica Kragic. "Social Robotics – Trust, Learning and Social Interaction." *Science Robotics* 13, no. 21 (2018): 1–2. <https://www.science.org/doi/epdf/10.1126/scirobotics.aau8839>.
- Hughes, James. *Citizen Cyborg: Why Democratic Society Must Respond to the Redesigned Human of the Future*. New York: Basic Books, 2004.
- Johnson, Deborah E., and Merel E. Noorman. "Responsibility Practices in Robotic Warfare." *Military Review*, May-June (2014): 12–21.

- https://www.armyupress.army.mil/Portals/7/military-review/Archives/English/MilitaryReview_20140630_art006.pdf.
- Kamiński, Stanisław, and Bogdan Czupryn. "Autonomia." In *Powszechna encyklopedia filozofii*. Vol. 1. Edited by Andrzej Maryniarczyk. Lublin: Polskie Towarzystwo Tomasza z Akwinu, 2000.
- Kanda, Takayuki, and Hiroshi Ishiguro. *Human-Robot Interaction in Social Robotics*. Boca Raton: CRC Press Taylor and Francis Group, 2013.
- Kisielewicz, Andrzej. *Sztuczna inteligencja i logika*. Warszawa: Wydawnictwo Naukowe PWN, 2017.
- Klichowski, Michał. *Narodziny cyborgizacji. Nowa eugenika, transhumanizm i zmierzch edukacji*. Poznań: Wydawnictwo Naukowe UAM, 2014.
- Kopeć, Rafał. "Autonomia systemów bojowych." *Przegląd Geopolityczny*, no. 17 (2016): 133–47.
- Krzysztofek, Kazimierz. "Sprawczość ludzka, transludzka i postludzka w społeczeństwie nasyconym technologicznie". In *Moc sprawcza ludzi i organizacji*. Edited by Lech Wojciech Zacher. Warszawa: Wydawnictwo Poltext, 2016.
- Kurzweil, Ray. *The Age of Spiritual Machine*. New York: Books Penguin, 1991.
- Lambert, Paul. "Computer Generated Works and Copyright: Selfies, Traps, Robots, AI and Machine Learning." *European Intellectual Property Review*, no. 39 (2017): 12–20.
- Latour, Bruno. *Reassembling the Social: An Introduction to Actor-Network-Theory*. New York: Oxford University Press, 2005.
- Lee, Kwan, Min, Namkee Park, and Hayeon Song. "Can a Robot be Perceived as a Developing Creature: Effect of a Robot's Long-Term Cognitive Developments on its Social Presence and People's Social Responses Toward it." *Human Communication Research*, no. 31 (2005): 538–63.
- Lem, Stanisław. *Bajki robotów*. Warszawa: Biblioteka Gazety Wyborczej, 2008.
- Loh, Janina. "Responsibility and Robot Ethics: A Critical Overview." *Philosophies* 4, no. 58 (2019): 1–20.
- Marx, Johannes, and Christine Tiefensee. "Of Animals, Robots and Men." *Historical Social Research* 40, no. 4 (2015): 70–91.
- Menzel, Peter, and Faith D'Aluisio. *Robo Sapiens: Evolution of a New Species*. Cambridge: MIT Press, 2001.
- Michalski, Ryszard. "O naturze uczenia się – problemy i kierunki badawcze." *Informatyka*, no. 2 (1988): 4–7.
- Miernicki, Martin, and Irene Ng [Ying Huang]. "Artificial Intelligence and Moral Rights." *AI & Society*, no. 36 (2020): 319–29.
- Mileszyk, Natalia. *Copyright trolling – skala zjawiska i propozycje rozwiązań*. Centrum Cyfrowe. <https://centrumcyfrowe.pl/czytelnia/copyright-trolling-skala-zjawiska-i-propozycje-rozwiazan/>.
- Misselhorn Catrin, *Robots as moral agents*. In *Ethics in Science and Society: German and Japanese Views*, edited by F. Rovekamp, F. Bosse., Monachium: Iudicium Verlag 2013.

- Mori, Masahiro. "The Uncanny Valley." *Energy* 7, no. 4 (1970): 33–5.
- Parasuraman, Raja, Thomas Sheridan, and Christopher Wickens. "A Model for Types and Levels of Human Interaction with Automation." *IEEE Transactions on Systems, Man and Cybernetics – Part A: Systems and Humans* 36, no. 3 (2000): 286–97.
- Pawińska, Maria. "Dopuszczalność gwałtu na humanoidalnym robocie." *Acta Universitatis Lodzensis. Folia Iuridica*, no. 86 (2019): 9–18.
- Przegalińska, Aleksandra. *Istoty wirtualne. Jak fenomenologia zmieniała sztuczną inteligencję*. Kraków: Wydawnictwo Universitas, 2016.
- Sarowski, Łukasz. *Robot humanoidalny jako podmiot czy przedmiot interakcji społecznej?* Lublin: XI Polski Zjazd Filozoficzny, 2019 r.
- Schmidt, Colin, and Felicitas Kraemer. "Robots, Dennet and the Autonomous: a Terminological Investigation." *Minds and Machines* 16, no. 1 (2006): 73–80.
- Schmidt, Colin, and Felicitas Kraemer. "Human and Computer Control of Undersea Teleoperations, Technical Report on Contract N00014-77-0256." https://www.researchgate.net/publication/23882567_Human_and_Computer_Control_of_Undersea_Teleoperators.
- Sheridan, Thomas B., and W. L. Verplank. *Human and Computer Control of Undersea Teleoperations, Technical Report on Contract N00014-77-0256*. https://www.researchgate.net/publication/23882567_Human_and_Computer_Control_of_Undersea_Teleoperators.
- Søraa, Roger. "Mechanical Genders: How do Humans Gender Robots." *Gender, Technology and Development* 21, no. 1–2 (2017): 99–115.
- Steels, Luc. *Social Learning and Verbal Communication with Humanoid Robots*. https://www.researchgate.net/publication/228800852_Social_learning_and_verbal_communication_with_humanoid_robots.
- Sullins, John. "When is a Robot a Moral Agent?" *International Review of Information Ethics* 6, no. 12 (2006): 23–30.
- Szpunar, Magdalena. *Kultura algorytmów*. Kraków: Instytut Dziennikarstwa, Mediów i Komunikacji Społecznej Uniwersytetu Jagiellońskiego, 2019.
- Szulczewski, Grzegorz. "Sztuczna inteligencja a inteligencja moralna. Zagadnienia wstępne cybernetyki." *Annales. Ethics in Economic Life* 22, no. 3 (2019): 19–31.
- Thurman, Neil. "Computational Journalism." In *The Handbook of Journalism Studies*. Edited by Karin Wahl-Jorgensen, and Thomas Hanitzsch. New York: Routledge, 2019.
- Veruggio, Gianmarco. "The Birth of Roboethics." In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2005): Workshop on Robo-Ethics, Barcelona, April 18, 2005*. <https://philpapers.org/rec/VERTBO-3>.
- Wierchoń, Michał, and Marta Łukowska. "Ucieleśnione poznanie." In *Przewodnik po kognitywistyce*. Edited by Józef Bremer. Kraków: Wydawnictwo WAM, 2016.
- Ziemke, Tom. "Czym jest to, co zwiemy ucieleśnieniem?" *Avant* 6, no. 3 (2015): 161–74.

ABSTRAKT / ABSTRACT

Łukasz SAROWSKI, Małgorzata GRUCHOŁA – Robotyka społeczna w perspektywie (nie)antropocentryzmu

DOI 10.12887/36-2023-3-143-10

Obecnie jesteśmy świadkami procesu osvajania sztucznej inteligencji w postaci, między innymi, tak zwanych robotów humanoidalnych towarzyszących – procesowi, który może doprowadzać do pewnego rodzaju ich „uspołecznienia”. Celem artykułu było wykazanie – poprzez analizę literatury przedmiotu oraz wytworów robotyki społecznej – że ów proces osvajania wymusza opracowanie uwzględniających założenia antropocentryzmu kryteriów i zasad funkcjonowania robotów w życiu społecznym. Analizie poddano cztery wymiary antropocentryzmu (poznawczy, ontologiczny, metodologiczny oraz aksjologiczny) w kontekście konieczności doprecyzowania przepisów etycznych, społecznych i prawnych wykorzystania robotów humanoidalnych w społeczeństwie. Zastosowano metody analityczno-opisową i syntetyczną. Uzyskane wnioski pozwalają na częściowe potwierdzenie przyjętej we wprowadzeniu hipotezy badawczej zakładającej, że proces osvajania technologii AI w postaci robotów humanoidalnych implikuje konieczność ich „uspołecznienia” we wszystkich czterech wymiarach antropocentryzmu. Z analizy literatury przedmiotu wynika, że rozwój robotyki społecznej zmienia sposób postrzegania i analizowania zjawisk społecznych, w których coraz większą rolę pełnią zrobotyzowane technologie. Realizują one zadania do niedawna zarezerwowane tylko dla człowieka, pełniąc różne funkcje społeczne. I chociaż na płaszczyźnie ontologicznej nie zostały jeszcze wypracowane formalne kryteria warunkujące podmiotowość robotów i nie przyznano im podmiotowości prawnej, to intensywny rozwój zrobotyzowanych technologii, zmieniając ich pozycję w przestrzeni społecznej, sprzyja nieformalnemu, uznaniowemu przyznaniu im podmiotowości. Zasadne wydaje się stwierdzenie, że przyznanie podmiotowości prawnej robotom humanoidalnym będzie generować zmiany we wszystkich obszarach antropocentryzmu.

Słowa kluczowe: robotyka społeczna, technologia sztucznej inteligencji, kryteria antropocentryzmu

Kontakt: (Łukasz Sarowski) Katedra Języka, Retoryki i Prawa Mediów, Instytut Dziennikarstwa i Zarządzania, Wydział Nauk Społecznych, Katolicki Uniwersytet Lubelski Jana Pawła II, Al. Raławickie 14, 20-950 Lublin; (Małgorzata Gruchola) Katedra Komunikacji Wizualnej i Nowych Mediów, Instytut Dziennikarstwa i Zarządzania, Katolicki Uniwersytet Lubelski Jana Pawła II, Al. Raławickie 14, 20-950 Lublin

E-mail: (Łukasz Sarowski) lukasz.sarowski@kul.pl; (Małgorzata Gruchola) malgorzata.gruchola@kul.pl

Tel. (Małgorzata Gruchola) 81 4455304

ORCID: (Małgorzata Gruchola) 0000-0002-2367-0416

Łukasz SAROWSKI, Małgorzata GRUCHOŁA, Social Robotics in the Perspective of (Non)Anthropocentrism

DOI 10.12887/36-2023-3-143-10

We currently witness a process of domesticating artificial intelligence in the form of, among others, the so-called humanoid companion robots; the process in question may lead to a kind of “socialization” of such robots. The purpose of the article was—through an analysis of the specialist literature and products of social robotics—to show that such domesticating makes us aware of the necessity to develop, considering anthropocentric principles, criteria and rules to regulate the functioning of robots in social life. Four dimensions of anthropocentrism (cognitive, ontological, methodological, and axiological) were analyzed in the context of the need to clarify ethical, social, and legal rules for the use of humanoid robots in society. The author applied analytical-descriptive and synthetic methods. The conclusions make it possible to partly confirm the research hypothesis, i.e., the claim that the process of domesticating AI technology in the form of humanoid robots implies the need for their ‘socialization’ in all four dimensions of anthropocentrism. The literature analysis shows that the development of social robotics is changing the perception and ways of analysis of social phenomena in which robotic technologies play an increasingly important role. Robots perform social functions and tasks which have so far been the domain of humans. Although from an ontological perspective, the formal criteria of subjectivity for robots have not been developed and robots have not been granted legal subjectivity, the intensive development of robotic technologies, by changing the position of robots in society, favours informal and discretionary granting them subjectivity. It seems justifiable to say that granting legal subjectivity to humanoid robots will generate changes in all areas of anthropocentrism.

Keywords: social robotics, artificial intelligence technology, criteria of anthropocentrism

Contact: Department of Language, Rhetoric, and Media Law, Institute of Journalism and Management, John Paul II Catholic University of Lublin, Al. Racławickie 14, Lublin, Poland

E-mail: (Łukasz Sarowski) lukasz.sarowski@kul.pl; (Małgorzata Gruchola) malgorzata.gruchola@kul.pl

Phone (Małgorzata Gruchola) +48 81 4455304

ORCID: (Małgorzata Gruchola) ORCID: 0000-0002-2367-0416

Anna MAJ

EWOLUCJA INTELIGENCJI I JEJ BADANIE Oswajanie sztucznej inteligencji w perspektywie przemian kultury i komunikacji

Roboty wyposażone w sztuczną inteligencję, które w komunikacji z człowiekiem przejawiają autoironię i działania irracjonalne, postrzegane są przez ludzi jako bardziej ludzkie i budzą pozytywniejsze odczucia. Sugerując, że także (jak ludzie) popełniają błędy, realizują pewien algorytm, dzięki któremu w ludzkim odbiorze stają się bardziej do nas podobne. Taka symulowana niekonsekwencja i niedoskonałość byłyby zatem tymi elementami, które sprawiają, że technologia przestaje budzić lęk i wydaje się bliższa innym doświadczeniom komunikacyjnym.

OSWAJANIE SZTUCZNEJ INTELIGENCJI JAKO PROCES KULTUROWY

„Oswoić” znaczy: przemyśleć, udomowić, nauczyć się komunikować, wreszcie przywyknąć. Proces udomowienia sztucznej inteligencji już trwa. Wciąż uczymy się z nią komunikować, jednak zdążyliśmy do niej przywyknąć. Nie zauważamy jej, choć towarzyszy nam prawie wszędzie, subtelnie i „bezszwowo” wnikając w naszą codzienność, zgodnie z ideą ubicomp i calm technologies¹. Elon Musk, wynalazca, przedsiębiorca i propagator AI, zwraca uwagę, że musimy przemyśleć, w którym kierunku będzie się ona rozwijać². Zaproponowane przez niego wbudowanie w proces badań nad sztuczną inte-

¹ Na idei calm technologies i paradygmacie wszechobecnego przetwarzania danych (ubiquitous computing, ubicomp) opierają się współczesne urządzenia komunikacyjne i ich interfejsy: komputery, laptopy, smartfony, interaktywne tablice, urządzenia zarządzające inteligentnym domem (ang. smart home) i miastem (ang. smart city) oraz wszystkie urządzenia i sensory związane z Internetem Rzeczy (ang. Internet of Things). Wdrażane dziś koncepcje przeprojektowania interfejsów nowych technologii tak, by stały się one przyjazne dla użytkownika, powstały w laboratoriach Xerox PARC u progu lat dziewięćdziesiątych (por. M. W e i s e r, *The Computer for the 21st Century*, „Scientific American” 256(1991) nr 3, s. 94-104; M. W e i s e r, J.S. B r o w n, *Designing Calm Technology*, Xerox PARC, 21.12.1996, <https://people.csail.mit.edu/rudolph/Teaching/weiser.pdf>).

² Po lekturze *Superinteligencji* Nicka Bostroma w roku 2014 Musk, twórca PayPal, Tesli, SpaceX i innych innowacyjnych przedsięwzięć, wyraził przekonanie, że sztuczna inteligencja może być groźniejsza od bomby atomowej. Z jego inspiracji powstała wówczas inicjatywa Open AI oraz Future of Life Institute. Open AI opisuje swoje cele jako prowadzenie badań i wdrożeń w obszarze AI, której „misją jest zapewnienie, że ogólna sztuczna inteligencja przyniesie korzyści całej ludzkości” (Open AI, *About*, <https://openai.com/about/>). Por. N. B o s t r o m, *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press, Oxford 2014 [wydanie polskie: *Superinteligencja. Scenariusze, strategie, zagrożenia*, tłum. D. Konowrocka-Sawa, Wydawnictwo Helion, Gliwice 2016]; E. M u s k, „Worth reading *Superintelligence* by Bostrom. We need to be super careful

ligencją etyczno-kulturowych hamulców w postaci inicjatywy Open AI nie musi jednak zakończyć się sukcesem. Bardziej prawdopodobne jest, że stanie się katalizatorem nieuchronnych zmian. Sukcesy chatbotów Open AI (GPT2, GPT3, a następnie ChatGPT3.5³) wskazują na to, że ów „etyczny projekt AI” stanie się kolejnym akceleratorem wielu gałęzi biznesu i dokona redefinicji kultury w obszarach tak istotnych, jak: edukacja (na wszystkich poziomach), przetwarzanie tekstu, dziennikarstwo, marketing, badania naukowe⁴, ekspertyza i szkolenia, programowanie, wyszukiwanie informacji i inne usługi związane z dostępem do wiedzy⁵. Oznacza to uwikłanie „projektu etycznego AI” w niezwykle zacięty wyścig o pierwszeństwo między sieciowymi gigantami, zwanymi „cytadelami danych” (Google, Microsoft, Amazon, Meta, Facebook), które nie bez powodu zwróciły się ku sztucznej inteligencji, zwłaszcza ku przetwarzaniu języka naturalnego za pomocą dużych modeli językowych (ang. Large Language Models)⁶. Przyjęcie konkretnych zasad etycznych w projektowaniu AI staje się zatem decyzją znaczącą dla kształtu cywilizacji przyszłości.

with AI. Potentially more dangerous than nukes”, Twitter, 3.08.2014, <https://twitter.com/elonmusk/status/495759307346952192>.

³ Pierwsza wersja API GPT3 została udostępniona do testów przez Open AI z pewnymi ograniczeniami 11 czerwca 2020, a następnie bez ograniczeń 18 listopada 2021. Na jego bazie stworzono następnie GPT-3.5 oraz ChatGPT-3.5, który udostępniono bez ograniczeń 30 listopada 2022, w ciągu pierwszego miesiąca przetestowało go z powodzeniem ponad milion użytkowników. T. C h a m o r r o - P r e m u z i c, *How ChatGPT Is Redefining Human Expertise: Or How To Be Smart When AI Is Smarter Than You*, „Forbes”, 12.01.2023, <http://bit.ly/3Ef5PUm>.

⁴ W czerwcu 2022 roku A.O. Thunström zadała GPT3 napisanie autotematycznego artykułu naukowego. Proces ten oraz jego konsekwencje w aspekcie psychologicznym i kulturowym został opisany przez badaczkę w „Scientific American”. Artykuł napisany przez AI został podpisany jego imieniem (GPT3 figuruje jako główny autor, a Thunström jako współautorka) i zgłoszony do recenzji w czasopiśmie naukowym, jest dostępny online w formie preprintu we francuskim archiwum HAL OA. Eksperyment Thunström przenosi rozważania nad autorstwem sztucznej inteligencji w świecie nauki na inny poziom, dotychczas wskazywano raczej na negatywne i pozytywne aspekty praktyk wspomaganie się GPT3 w pisaniu prac naukowych, lecz nie na możliwość wykazywania AI jako współtwórcy tekstów (badań). Por. G p t G e n e r a t i v e P r e t r a i n e d T r a n s f o r m e r, A.O. T h u n s t r ö m, S. S t e i n g r i m s s o n, *Can GPT-3 write an academic paper on itself, with minimal human input?*, HAL [Open Science Archive], <https://hal.science/hal-03701250>; A.O. T h u n s t r ö m, *We Asked GPT-3 to Write an Academic Paper about Itself – Then We Tried to Get It Published*, „Scientific American”, 30.06.2022, <http://bit.ly/3IDS4YM>.

⁵ S. M i n t z, *ChatGPT: Threat or Menace?*, „Inside Higher Ed”, 16 I 2023, <http://bit.ly/412ObwI>.

⁶ Metaforę „cytadeli danych” zaproponował Martin Warnke (por. M. W a r n k e, *Databases as Citadels in The Web 2.0*, w: *UnLike Us Reader. Social Media Monopolies and Their Alternatives*, red. G. Lovink, M. Rasch, Institute of Network Cultures, Amsterdam 2013, s. 86). Projekty wielkich korporacji IT w obszarze AI reprezentują między innymi: ChatGPT-3 (Open AI we współpracy z Microsoftem, 2022), LaMDA (Google, 2021), BlenderBot3 / OPT-175B (Facebook / Meta, 2022), Alexa (Amazon, 2018).

Warto przyrzeć się nie tylko w społecznym przejawom „oswajania inteligencji”, ale też bardziej nieoczywistemu procesowi ewolucji inteligencji (od biologicznej, przez sztuczną, porozszerzoną) i dotychczasowym interdyscyplinarnym badaniom nad sztuczną inteligencją, na przykład rozwojowi sposobów rozumienia terminu „inteligencja” w tym obszarze. Z uwagi na szerokość tego typu badań przedstawione zostaną tu jedynie wybrane koncepcje i eksperymenty, pochodzące z różnych obszarów wiedzy oraz z pogranicza nauki i sztuki nowych mediów. Wieloletnie i interdyscyplinarne badania nad sztuczną inteligencją (a także cyborgizacją, robotyzacją, poznaniem i percepcją) przyczyniły się do wypracowania koncepcji i rozwiązań technicznych, które wpływają pośrednio na cyberkulturowe przemiany konceptualizacji ciała i umysłu oraz na współczesne praktyki kulturowe, zwłaszcza w zakresie zachowań komunikacyjnych między innymi w sferze relacji międzyludzkich czy transformacji znaczenia przestrzeni domowej i publicznej.

Wysokie wartościowanie w cyberkulturze kategorii nowości i rozwoju technologicznego oraz atrakcyjność nowych interfejsów i funkcjonalności urządzeń sprzyjają wytworzeniu się i popularyzacji protechnologicznego dyskursu, który zachęca potencjalnych użytkowników nowych technologii do zaangażowania i interakcji z pomocnymi, inteligentnymi aplikacjami, agentami konwersacyjnymi⁷ i robotami. Prowadzi to do humanizacji technologii, a jednocześnie do pozytywnego wartościowania maszyny per se i postrzegania cyborgizacji jako naturalnego i dobroczynnego procesu.

⁷ W naukowej i branżowej literaturze przedmiotu stosowane są wymiennie terminy: „agent konwersacyjny”, „inteligentny agent wirtualny”, „agent sztucznej inteligencji”, „konwersacyjna sztuczna inteligencja (AI)”, „chat bot”, „bot konwersacyjny”, „asystent komunikacyjny”, „asystent głosowy”, „cyfrowy asystent/ka”, „aplikacja sztucznej inteligencji” i tym podobne. Opisują one oprogramowanie działające w określonym celu i wykorzystujące sztuczną inteligencję, umożliwiające użytkownikowi naturalną komunikację za pośrednictwem konwersacji (najczęściej głosowej, ale także tekstowej). Ich istotą jest pozyskiwanie i przetwarzanie danych nie tylko z predefiniowanej i zamkniętej bazy danych, ale też na podstawie danych pozyskanych z sieci, wielkich zbiorów danych oraz wieloetapowego procesu „uczenia się” na podstawie interakcji z użytkownikiem (ang. machine learning), które zazwyczaj trwa wiele miesięcy. Przykładami takiego oprogramowania są między innymi: Apple Siri, Google Now, Google Assistant, Microsoft Cortana, Amazon Alexa. Przykład definicji z obszaru literatury branżowej por. Microsoft Learn, *Co to są agenci sztucznej inteligencji?*, 01.12.2022, <https://learn.microsoft.com/pl-pl/azure/cloud-adoption-framework/innovate/best-practices/conversational-ai>; z kolei wczesne naukowe opracowanie kwestii terminologii por. A. J a n a s, *Zastosowanie agentów konwersacyjnych w nauczaniu*, „Nierówności Społeczne a Wzrost Gospodarczy” 2011, nr 22, s. 281. Termin „agent konwersacyjny”, jakkolwiek wydaje się niezgrabny językowo, jest jednym z częściej stosowanych przez naukowców, projektantów i producentów oprogramowania.

UMYSŁ, CIAŁO, KOMUNIKACJA

Transhumanizm i cyborgizacja, zwłaszcza koncepcja „upgradeowania” ludzkiego ciała i umysłu, dawno już przeszły od futurologicznych wizji nie tylko do naukowych (w tym medycznych) praktyk, ale i do popkulturowych realizacji⁸. Na poziomie rozszerzania możliwości umysłu w ciągu ostatniej dekady dokonała się znacząca zmiana. Prawie wszyscy przechodnie na ulicach współczesnych miast i w przestrzeniach publicznych pochłonięci są nie rozmowami z innymi ludźmi, ale zapatrzeni w ekrany smartfonów, tabletów i laptopów; delegują swe umysły do cyberprzestrzeni, zatapiając się w wirtualnym kontakcie z nieobecnymi lub poszukując informacji w sieci. Doświadczeniem komunikacyjnym, pogłębionym przez pandemię i rozwój usług sieciowych, stała się rozmowa nie z drugim człowiekiem, a z chatbotem, czyli agentem konwersacyjnym opartym na sztucznej inteligencji, takim jak Siri czy Alexa. Wykorzystywany przez tego typu oprogramowanie interfejs głosowy (zwany „konwersacyjnym interfejsem użytkownika”, CUI) symuluje do pewnego stopnia naturalną komunikację międzyludzką dzięki przetwarzaniu i modelowaniu języka naturalnego. Na razie trudno ocenić, czy zachodzi tu realny upgrade, czyli zwiększenie możliwości intelektualnych człowieka, czy częściej downgrade, dewaluacja ludzkich zdolności kognitywnych, odpowiadająca ewolucyjnie wypracowanym potrzebom szybkiej gratyfikacji, jaką daje dostęp do niekończącego się strumienia informacji⁹. Kultura partycypacji

⁸ Zgodnie z deklaracją transhumanistyczną wypracowaną przez World Transhumanist Association (1998) na podejście transhumanistyczne składają się: idea transcendencji biologii oraz przyspieszenia rozwoju człowieka za pomocą wszelkich dostępnych narzędzi technologicznych (pragmatyzm), a także memetyczna propagacja idei transhumanizmu i protechnologicznego dyskursu, idea realizacji celów transhumanizmu we wszystkich obszarach życia społecznego, uznania dla różnorodności koncepcji i idea nieustannej ewolucji w zgodzie z rozwojem technologicznym transludzkości (zob. The World Transhumanist Association. H+pedia [wiki]: *Transhumanist Declaration* [hasło], https://hpluspedia.org/wiki/Transhumanist_Declaration. Kontynuacją aktywności stowarzyszenia jest działalność Humanity+). Organizacja w swoim serwisie deklaruje przywiązanie do idei etycznego wykorzystania technologii, ale też do promowania badań i wdrożeń technologii przyszłości wykraczających poza dzisiejszą „normalną” ludzką egzystencję, to jest przedłużanie życia i poszerzenie możliwości ciała i umysłu poprzez technologiczne i biotechnologiczne protezy, bioniczne rozszerzenia i terapie genowe (zob. Humanity+, <https://www.humanityplus.org/about>). Opisywaną w tym kontekście cyborgizację można natomiast zdefiniować jako „wspomaganie funkcji umysłowych i życiowych zdrowych ludzi” (P. Łaszczyca, *Człowiek i jego maszyny. Operatorzy i protezy*, „Filo-Sofija” 17(2017) nr 39(1), s. 58). O popkulturowych realizacjach pisze: E. Olzacka, *Od Homo sapiens do Homo immortalis. Idea nieśmiertelności w rosyjskich projektach filozoficznych i społeczno-politycznych*, w: *Technokultura: transhumanizm i sztuka cyfrowa*, red. D. Gałuszka i in., Wydawnictwo Libron, Kraków 2016, s. 47-62.

⁹ Badania na temat neurologicznych podstaw zachowań związanych z konsumpcją informacji, zarówno użytecznej, jak i bezużytecznej, wskazują, że informacja sama w sobie „postrzegana” jest przez mózg jako nagroda. Tłumaczy to zachowania związane z kompulsywnym korzystaniem

i konsumpcji symbolicznej, wirusowe rozsiewanie treści za pomocą sieci społecznościowych i narzędzia dające poczucie aktywności oraz relacji, sprzyjają transformacji umysłu człowieka niekoniecznie (lub nie zawsze) w kierunku jego doskonalenia¹⁰.

Przemiana dokonuje się także na poziomie konceptualizacji ciała¹¹. Liczne przeróbki, tak zwane „remake’i” ciała można postrzegać jako pierwszy stopień akceptacji procesu cyborgizacji¹². Ciało jest traktowane jako przedmiot

z mediów cyfrowych, w tym infoholizm (por. K. Kobayashi, M. Hsu, *Common Neural Code for Reward and Information Value*, „PNAS” 116(2019) nr 26, s. 13061-13066, <https://www.pnas.org/doi/full/10.1073/pnas.1820145116>).

¹⁰ Niewątpliwie, towarzyszące korzystaniu z mediów i z usieciowionej informacji, poczucie omnipotencji buduje jednak we współczesnym człowieku wrażenie dostępności wiedzy, która nie została zinternalizowana, a w konsekwencji – ograniczenie procesu uczenia się. Wiedza dostępna potencjalnie dzięki asyistentom AI nie jest jednak wiedzą przetworzoną i poddaną refleksji w umyśle człowieka, jest to wiedza technicznie zapośredniczona, zdalna i peryferyczna.

¹¹ Popkulturowa wizja ciała to ciało wytatuowane, zmodyfikowane, urozmaicone – poprzez różnego rodzaju ozdoby, obrazy, implanty. Trend ten zapoczątkowały nie tylko ruchy kontrkulturowe, ale też sztuka nowych mediów, między innymi dzieła Stelarc’a (robotyczne nogi i ręce, egzoskielety, wreszcie implant ucha na przedramieniu). Artysta przeszedł drogę od maltretowania własnego ciała w trakcie performansów z podwieszaniem i nakłuwaniem go, przez bolesne eksperymenty robotyczne, podpinające jego ciało do robotycznych protez i sieci (wraz z oddaniem nad nim pełnej kontroli publiczności), po radykalny eksperyment, w koncepcji trwale modyfikujący jego ciało poprzez wszczepienie implantu ucha na rękę, czyli sztucznego organu komunikacyjnego, który miał pozwalać nasłuchiwać artyście treści z internetu. Ostatnie dzieło można uznać za nie do końca udane, gdyż implant nie zadziałał i trzeba było go usunąć – zbuntowało się bowiem ciało Artysty. Dokumentacja prac por. *Suspensions, Third Hand, Exoskeleton, Muscle Machine, Ear on Arm*, Stelarc, <http://stelarc.org>. Działalność takich artystów należy postrzegać jako świadomy dialog człowieka z technologią, stymulujący przesuwanie granic „wyobraźnialnego”. Eksperymenty te spotykały się zawsze z dużym oddźwiękiem zarówno wśród krytyków sztuki mediów, jak i szerokiej publiczności. Popkultura przetwarza takie bodźce docierające ze świata sztuki, odzierając je z sensów i znaczeń dyskursotwórczych. Pozostają znaki bez oznaczników, puste ornamenty. Zamiast poważnego dyskursu na temat przyszłego rozwoju człowieka, w świadomości społecznej pozostaje obraz ucha na rękę i towarzyszące mu zadziwienie oraz akceptacja biohybrydyzacji. Ciało naturalne (w znaczeniu niezmodyfikowane), podobnie jak niezmediatyzowany umysł, przestaje mieć rację bytu – a w każdym razie części społeczeństwa wydaje się nudne. Kultura pielęgnuje dziś model „DIY: do it yourself”. Kreatywność przenoszona jest na otoczenie, a tym samym również ciało staje się jej nośnikiem i łatwo dostępnym medium.

¹² Kolejnym jest afirmacja biohybrydyzacji, w której wyraźną pomocą staje się chirurgia plastyczna, niestroniąca od kreatywnych realizacji, pozwalających na „transgatunkowe” czy „intergalaktyczne” przeróbki twarzy i ciała klientów (na przykład zamianę w wybranego bohatera komiksu czy ulubionej serii science-fiction). W tym kontekście interesujące wydają się uwagi Anny Irsak na temat przemiany etyki w obrębie współczesnej medycyny, a może nawet wizji całej dyscypliny i postrzegania jej zasadniczych celów (koncepcja kontraktualistyczna). Konstatacje te w znacznej mierze dotyczą także istoty i zasadniczych celów medycyny estetycznej, co uwiadcza się zwłaszcza w obrębie jej „artystycznych”, nie interwencyjnych (to jest niemających na celu pomocy medycznej) zastosowań podszytych pogonią za sławą i oryginalnością, a często ideologią transhumanistyczną (por. A. Irsak, *W obronie (tradycyjnej) medycyny*, „Ethos” 28(2015) nr 3(111), s. 176-188).

i środek przekazu¹³. Zakorzenie umysłu w technologii jest jednak znacznie bardziej powszechne i uzależniające, niesie też poważniejsze konsekwencje na przyszłość dla człowieka i kultury. Smartfony, wyposażone w agenta konwersacyjnego AI i potencjalnie nieskończenie rozszerzalną funkcjonalność (za pośrednictwem aplikacji i sklepów z aplikacjami), stają się prawdziwymi czynnikami zmiany kulturowej.

Istnieją już dowody na to, że małe dzieci uczące się mówić, posiadające w domu konwersacyjną sztuczną inteligencję, komunikują się inaczej, tworząc nowy rodzaj języka¹⁴. Dzieci adaptują się nieświadomie do sprawniej komunikacji z AI, dostosowując struktury językowe do wydawania poleceń „maszynie” i uzyskiwania od niej odpowiedzi. Struktury gramatyczne w takiej konwersacji różnią się od tradycyjnej (wyłącznie ludzkiej) socjalizacji, co więcej, zostają przeniesione na komunikację z rodziną. Agent konwersacyjny nie staje się może kolejnym domownikiem, ale z pewnością zmienia kontekst komunikacyjny, a także powoduje modyfikację pewnych zachowań komunikacyjnych. Obecność sztucznej inteligencji w środowisku domowym, takiej jak choćby Alexa (Amazon Echo), nie tylko zmienia nastrój wnętrza (na przykład włączając muzykę), ale też – mniej dostrzegalnie – redefiniuje relacje międzyludzkie w rodzinie¹⁵ i oswaja sztuczną inteligencję jako „naturalny” (bo wszechobecny i „udomowiony”) element bliskiego środowiska (domu). Może też – i od pewnego czasu już to czyni – rozwiązywać problemy najwyższej wagi: ratować domownikom życie, wzywając pogotowie ratunkowe¹⁶. Konsekwencją takich zależności komunikacyjnych i egzystencjalnych jest wzrost zaufania wobec nowej technologii i delegowanie na nią ludzkiej odpowiedzialności. Skoro AI ratuje życie i doradza w kwestiach zdrowotnych, rozwiązuje zadania domowe, zastępując w procesie edukacji dzieci pomoc rodzica¹⁷, dostarcza

¹³ Medycyna plastyczna w takich realizacjach wydaje się spełniać jedynie funkcję pozbawionego refleksji etycznej narzędzia rynkowego. Można rozumieć wspomniane modyfikacje ciała jako znak pojawienia się nowych tożsamości, jak chcą niektórzy badacze, zdaniem innych obrazują one kryzys duchowości i szacunku do ciała (zob. „Człowiek a/i lub maszyna: Duchowość 3.0”, seminarium, Wydział Teologiczny Uniwersytetu Śląskiego w Katowicach, 13 kwietnia 2018; „Filo-Sofja” 17(2017) nr 39(1)).

¹⁴ Por. A. Hiniker i in., *Can Conversational Agents Change the Way Children Talk to People?*, „Interaction Design and Children”, ACM, June 24, 2021, s. 338-349.

¹⁵ Por. E. Beneteau i in., *Parenting with Alexa: Exploring the Introduction of Smart Speakers on Family Dynamics*, w: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ACM, New York, April 2020, s. 1-13.

¹⁶ Por. Z. Klemm i in., *Apple's Siri Calls Ambulance for Baby*, „BBC News”, 07.06.2016, <https://www.bbc.com/news/technology-36471180>.

¹⁷ Zob. A. Agraval i in., *Prediction Machines. The Simple Economics of Artificial Intelligence*, Harvard Business School Publishing, Boston 2018.

wszelkich informacji – nie tylko staje się ekspertem, który pomaga decydować, ale też człowiek zaczyna odczuwać wobec niego wdzięczność i przestaje ufać własnym kompetencjom.

To zapośredniczenie następuje na wielu poziomach komunikacji, w pewnym stopniu niezauważalnie. W korzystaniu z asystenta AI czy opartej na AI aplikacji nie ma nic podejrzanego, oferują one komodyfikację i racjonalizację naszych działań, zwiększając efektywność podejmowanych czynności i dając poczucie optymalizacji zachowań i wyborów. *Sztuczna inteligencja, oferując automatyzację zachowań i wyborów, zastępuje coraz częściej człowieka w podejmowaniu drobnych, codziennych decyzji.* Podpowiada, jakie ubrania w naszej szafie lepiej do siebie pasują, co powinniśmy dziś zjeść i co kupić, jak dojechać najsprawniej do danej lokalizacji, z kim powinniśmy się spotykać, jakie wspomnienia celebrować, a także kiedy iść spać, by wypaść się najlepiej¹⁸. Algorytmizacja codzienności oraz zbieranie wzorców mikrobehawioralnych następują niepostrzeżenie, przekraczając kolejne granice prywatności.

Transhumaniści mówią o potrzebie integracji nośników cyfrowych danych z ciałem (za pośrednictwem chipów komunikacyjnych lub implantów domózgowych i nanotechnologii), lecz nie wydaje się to konieczne. Współczesny etap „kultury kciuka” i „szyi smartfonowej” oznacza wystarczającą integrację z technologią, optymalną z perspektywy przystosowania ewolucyjnego ciała i aparatu percepcyjnego. Używając inteligentnego smartfona, *homo sapiens* korzysta bowiem, jak dotąd, to jest w całym procesie ewolucji, z dostępnych w środowisku narzędzi (używał ich już przecież *homo habilis*, czyli człowiek (już) zręczny, choć (jeszcze) nie mądry, to jest nie *sapiens*). Współczesne narzędzia komunikacyjne dają poczucie kontaktu, dostęp do informacji, a jednocześnie pozwalają delegować umysł i dostosować aparat percepcyjny do nowych wyzwań usieciowionej technocodzienności.

Ten etap ewolucyjny zostanie przekroczony, gdy tylko ciało człowieka dostosuje się trwale do narzędzi cyfrowych, a cyborgizacja będzie codziennością. Stanie się to wraz z przekroczeniem granic oporu ciała i kultury. To jedna z możliwych perspektyw spojrzenia na przełomowy punkt nadejścia „osobliwości” (ang. singularity), projektowany przez Raya Kurzweila, wynalazcę, futurologa i promotora transhumanizmu¹⁹. Transgresja jest przecież wpisana w transhumanistyczny proces postępu technologii i ewolucji człowieka. Wy-

¹⁸ Przykłady aplikacji realizujących powyższe zadania dzięki AI: Acloset – AI Fashion Assistant, FoodVision: Ai Food Tracker, Google Maps, FindBee: Find My Family&Friends, iPhoto/iCloud, Smart Self-Control to Focus, Smart Self-Control to Wake Up et cetera (Apple Store).

¹⁹ Por. R. Kurzweil, *The Singularity is Near: When Humans Transcend Biology*, Duckworth, London 2009, s. 260. Zob. też E. Nadchodzi osobliwość. Kiedy człowiek przekroczy granice biologii, tłum. E. Chodkowska, A. Nowosielska, Kurhaus Publishing, Warszawa 2016.

czekiwany moment osobliwości to narodziny superinteligencji (i tym samym kres człowieka biologicznego)²⁰.

CYBORGIZACJA CZŁOWIEKA I HUMANIZACJA MASZYN

Ted Nelson, genialny twórca koncepcji hipertekstu i filozof epoki cyfrowej, już w latach siedemdziesiątych przestrzegał w swym wpływowym manifestie *Computer Lib/Dream Machines* przed postrzeganiem komputerów i AI jako narzędzi myślących²¹. Badacz akcentował demiurgiczną rolę człowieka wobec technologii komputerowych. Transhumaniści dążą jednak do zmiany relacji między człowiekiem i technologią, widząc w cyborgizacji człowieka i humanizacji maszyny szansę. Zamiast patrzeć na transhumanizm jedynie z perspektywy filozoficznej, dostrzegając w nim propozycję na czasy kryzysu humanizmu²², można też postrzegać go z perspektywy kulturowej i ekonomicznej. To nowy rodzaj darwinizmu społecznego – darwinizm technologiczny, planowany dla przyszłych pokoleń przez współczesne. Dostęp do nowych technologii (biotechnologii, nanotechnologii, protetyki, augmentacji) będzie miał decydujący wpływ na pozycję jednostki w społeczeństwie: poprzez ciało i umysł augmentowane według woli i możliwości finansowych rodziców przyszłych dzieci²³. Sztuczna inteligencja będzie pełniła tu rolę podstawową i definiującą dla postspołeczeństw przyszłości, zarządzając nimi i optymalizując wszelkie wysiłki jednostek.

Współczesne podejście protechnologiczne, jednoczące wyraźnie świat nauki, polityki i globalnej gospodarki, w którym sztuczna inteligencja oraz roboty stanowią nadzieję ludzkości, zakorzenione jest w myśli transhumanistycznej. Marvin Lee Minsky, jeden z wiodących myślicieli transhumanizmu, pionier

²⁰ „Superinteligencja” rozumiana jest jako sztuczna inteligencja przekraczająca poziom ludzkiej inteligencji, co ma uniemożliwić konkurowanie człowieka z maszynami (por. I. J. G o o d, *Speculations Concerning the First Ultrainelligent Machine*, „Advances in Computers” 6(1965), s. 31-88).

²¹ Por. T. N e l s o n, *Computer Lib/Dream Machines*, <http://worrydream.com/refs/Nelson-ComputerLibDreamMachines1975.pdf>.

²² W zależności od ujęcia badaczy oznacza to: nowy humanizm (Carry Wolfe) lub antyhumanizm (posthumanizm), czyli albo odrodzenie myśli humanistycznej, lecz w nowym kontekście technologicznym (stawiającej człowieka i jego rozwój w centrum zainteresowania), albo zaprzeczenie wszystkim wartościom humanizmu (konstatacja końca epoki człowieka, akcentowanie destrukcyjnego dla planety i innych gatunków wpływu człowieka, zgoda na przejście świata przez nową formę, czyli maszyny). Por. M. G a r b o w s k i, *Transhumanizm: Geneza – założenia – krytyka*, „Ethos” 28(2015) nr 3(11), s. 34; E. B i ũ c z y k, *Inżynieria klimatu a inżynieria człowieka. Dyskursy na temat środowiska w epoce antropocenu*, „Ethos” 28(2015) nr 3(11), s. 156.

²³ Zob. F. F u k u y a m a, *Transhumanism. A Special Report*, „Foreign Policy”, 23.10.2009, <https://foreignpolicy.com/2009/10/23/transhumanism>.

sztucznej inteligencji i robotyki, był jednym z patronów instytucjonalizacji badań nad sztuczną inteligencją²⁴. To dzięki niemu oraz Johnowi McCarthy'emu powołano instytuty naukowe zajmujące się AI w Stanford i w MIT. Równolegle podobne badania rozpoczęły korporacje IBM, Rand czy Bell Labs²⁵. To w tych laboratoriach wypracowano nowe koncepcje inteligencji i ich przełomowe aplikacje w obszarze łączącym informatykę z robotyką oraz teorią poznania i percepcji. Pojęcie „inteligencji” nieodwracalnie zostało zredefiniowane, przestało być kategorią opisową zarezerwowaną dla działań człowieka. Wcześniejszego rozszerzenia zakresu pojęcia „inteligencji” na świat zwierząt dokonali cybernetycy wspólnie z etologami i badaczami systemów biologicznych. Dzisiejsze dyscypliny badań nad sztuczną inteligencją, robotyka, protetyka, informatyka i wiele innych korzystają z dokonań obydwu nurtów. Transhumanistyczne założenia o logice nieustannego rozwoju systemów oraz o prawie do pozabiologicznego przyspieszenia rozwoju człowieka upowszechniają się wraz z ich sukcesami²⁶.

Współcześnie powszechnie i bezdyskusyjnie zakłada się, że wszystkie kryzysy zniweluje jedynie sztuczna inteligencja, zgodnie z rozpropagowaną japońską koncepcją Społeczeństwa 5.0, według której wszelkie problemy społeczne mają być rozwiązywane przez nowe technologie²⁷. Oddelegowa-

²⁴ Na przykład Marvin Minsky jest twórcą pierwszej sieci neuronowej (SNARC), wynalazcą jednego z pierwszych skanerów wizualnych, pierwszych robotów – „żółwia” LOGO, robotycznej dłoni z czujnikami dotyku oraz wielu interfejsów sprzętowych i oprogramowania, uhonorowany został między innymi nagrodą Turinga. Zob. Brief Academic Biography of Marvin Minsky, MIT, <https://web.media.mit.edu/~minsky/minskybiog.html>.

²⁵ Por. H. R h e i n g o l d, *Narzędzia ułatwiające myślenie. Historia i przyszłość metod poszerzania możliwości umysłu*, tłum. J. Szporko, Wydawnictwa Naukowo-Techniczne, Warszawa 2003, s. 150. Niebagatelny wpływ na tę dyscyplinę mieli Claude Shannon, Herbert A. Simon czy Allen Newell. Istotną rolę w konceptualizacji obszaru badawczego odegrały zwłaszcza prekursorskie prace Alana Turinga oraz Shannona, zwracające uwagę na inne aspekty niż biologicznie zorientowana cybernetyka Norberta Wienera (por. A. M a j, *Przemiany wiedzy w cyberkulturze. Badania nad kulturą, komunikacją, wiedzą i mediami*, Wydawnictwo Uniwersytetu Śląskiego, Katowice 2021, s. 335).

²⁶ Pewną rolę w popularyzacji koncepcji sztucznej inteligencji w świadomości społecznej odegrały również powiązania Minsky'ego z Hollywood. Minsky oraz Irving J. Good byli doradcami Stanleya Kubricka w realizacji filmu *2001: Odyseja kosmiczna* (USA, 1968), który wpłynął na kolejne pokolenia twórców. Minsky miał znaczący wpływ na filmową wizję komputera HAL 9000, znacznie różniącą się od wersji z noweli Arthura C. Clarka (zob. D.G. S t o r k *Scientist on the Set: An Interview with Marvin Minsky*, w: *HAL's Legacy: 2001's Computer as Dream and Reality*, red. D.G. Stork, wstęp A.C. Clarke, MIT Press, Cambridge, London 1998). Kolejne sukcesy nowych technologii przyemiły jednak przez lata dystopijne wizje Kubricka i innych artystów, a technoop-tymistyczna polemika inżynierów i informatyków dominuje dziś nad bardziej krytycznym wobec technologii dyskursem humanistycznym.

²⁷ Por. J. G r u d o w s k a, D. Z i e l i Ń s k i, *Społeczeństwo 5.0: Refleksja krytyczna*, „Transformacje” 2(2022), s. 234. Wśród kryzysów rozwiązywalnych przez AI wymienia się w literaturze na przykład katastrofę klimatyczną, pandemię, epidemie, trudne operacje medyczne, problematyczne

nie odpowiedzialności i przekazanie jej systemom ekspertowym już jest codziennością w coraz większej ilości obszarów życia społecznego. Systemy opierające się na analizie big data stają się nieodzowne, zmieniając logikę współczesnego działania człowieka i pozbawiając go ciężaru decyzyjnego, który dotąd definiował kondycję ludzką. Maszyny i systemy informatyczne, a zatem narzędzia nowej generacji, stały się ośrodkiem zaufania jako reprezentanci porządku algorytmicznego, któremu od oświecenia zawierzamy życie. Od koncepcji mózgu elektronowego, popularnej w połowie dwudziestego wieku, przeszliśmy daleką drogę do autonomicznych i samouczących się sieci neuronowych, których przetwarzania informacji nie jesteśmy w stanie w pełni zrozumieć.

Na ten kierunek rozwoju społeczeństwa informacyjnego i stopniową algorytmizację kultury zwracali od lat uwagę badacze z różnych dyscyplin²⁸. Dziś dyskusja filozoficzna i humanistyczna podążać musi za nowymi przełomami w naukach informatycznych, biotechnologii i robotyce, tym samym jest ona sprowadzana do podwójnej roli – przestrzegania i komentowania.

Podwójną rolę początkowo pełnili także refleksyjni eksperymentatorzy i wynalazcy, tacy jak Kevin Warwick, profesor cybernetyki na University of Reading, a następnie University of Coventry w Wielkiej Brytanii, „pierwszy cyborg”, czy Steve Mann, profesor inżynierii elektrycznej i komputerowej z University of Toronto, „ojciec technologii ubieralnych”. Naukowcy i artyści zarazem, którzy u progu dwudziestego pierwszego wieku podejmowali pierwsze próby cyborgiczne na sobie samych, stawiali intrygujące pytania o dalsze możliwości współegzystencji człowieka i maszyny oraz o jej zasady. Ryzykując własnym zdrowiem, przesuwali granice poznania. Mann i Warwick niezależnie pracowali nad możliwością wykorzystania nowych technologii w protetyce i w obszarze assistive technologies, czyli technologii wspierających osoby niepełnosprawne. Ich dokonania wpłynęły na współczesną medycynę. Mann, eksperymentując w obszarze substitucji i technicznego rozszerzania zmysłu wzroku, rozwinął koncepcję technologii ubieralnych (ang. wearables). Warwick poddał się ope-

diagnozy, zarządzanie przestrzenią miejską czy infrastrukturą krytyczną, sprawiedliwy osąd więźniów lub rozpoznanie podejrzanych, wytypowanie pracowników do zwolnienia.

²⁸ Por. T. M i c k a, *Specifics of Algorithmization in Data Culture*, w: *ICSRS 2019: Conference Proceedings, International Conference on Recent Social Studies and Research*, 25-26 October 2019, CORIS Department Sapienza University, Rome 2019, s. 240-247; K. K r z y s z t o f e k, *Człowiek – społeczeństwo – technologie. Między humanizmem a transhumanizmem i posthumanizmem*, „Ethos” 28(2015) nr 3(111), s. 191-213; L.W. Z a c h e r, *Od społeczeństwa informacyjnego do społeczeństwa wiedzy (dylematy tranzycyjne: między informacją, wiedzą i wyobraźnią)*, w: *Spółeczeństwo informacyjne. Wizja czy rzeczywistość?*, t. 1, red. L.H. Haber, Akademia Górniczo-Hutnicza, Kraków 2004, s. 103-112; T. B i e l a k, G. P t a s z e k, *Algorytmiczne doświadczanie kultury*, „Kultura Współczesna” 2019, nr 1(104), s. 10-14.

racjom wszczęcia kolejno dwóch chipów w swoje ciało, czyli dobrowolnej cyborgizacji, i przeprowadził prekursorskie badania w zakresie transferu sygnału na odległość, zdalnej kontroli robotów i systemów telekomunikacyjnych oraz zachowań ludzkiego ciała i umysłu podłączonego do komputera²⁹. Badacz od dawna przekonuje, że jego doświadczenia i sukcesy komunikacyjne nowych technologii skazują tych, którzy nie będą chcieli się zmodyfikować poprzez biochipowanie i protezowanie ciała na ewolucyjną klęskę. Porównuje komunikację cyborgów i ludzi ze współczesną komunikacją (odpowiednio) ludzi i krów³⁰.

Być może perspektywa szybszego rozwoju bardziej sprawnych procesów myślowych zwycięży, jednak na razie obserwujemy raczej drugą stronę tego procesu: doskonalenie się sztucznej inteligencji, uczłowieczanie robotów i rozwój systemów ekspertowych zarządzających biologicznie nierozszerzonymi ludźmi i społeczeństwami. Warwick dopuszcza jednak ewolucję podejścia społeczeństw do kwestii upgrade'u, sam bowiem przeszedł od lęku do fascynacji możliwościami cyborga³¹.

Idee cyborgizacji ciała i sztucznej inteligencji spotykają się w usieciowieniu i dostępie do wszechobecnych danych (ang. ubicomp). Cyborg jawi się jako sprawniejszy mentalnie, a zarazem lepiej przetwarzający dane w czasie rzeczywistym. Dziś systemy ekspertowe, agenty konwersacyjne i różne systemy oparte na sztucznej inteligencji karmione są wielkimi zbiorami danych i trenowane na określonych modelach po to, by same się uczyły. Nawet metaforyka kontaktu ze sztuczną inteligencją przypomina model rodzicielskiego kontaktu z małym dzieckiem. Daje to poczucie kontroli, ale jednocześnie mówi o kreatywnym aspekcie tego procesu. Kreatywność kojarzona bywa zazwyczaj pozytywnie, dzięki czemu taki dyskurs może mieć charakter techniczny (skupia się na działaniach), zarazem nie wywołując lęku. Jednak po sukcesach sztucznej inteligencji w grze w go, gdy Google DeepMind pokonał Lee Sedolę³², oraz po raporcie o chatbotach wymyślających własny język, by samodzielnie negocjować, można mówić o nowych obszarach budzących potencjalnie lęk, ale też refleksję nad zachodzącymi procesami humanizacji technologii, robotów i AI. Dyskusja nad sukcesami „twórczego” i „rozumnego” GPT-3.5 oraz „wrażliwego” i „podmiotowego” LaMDA jest kolejną odsłoną tego procesu: GPT-3.5 przechodzi test Turinga, zdobywając uznanie naukowców i dzienni-

²⁹ Zob. K. Warwick, *I, Cyborg*, University of Illinois Press, Urbana–Chicago 2004.

³⁰ Por. Infomia, *Cyborg Life: Kevin Warwick*, YouTube, 14.04.2008, http://www.youtube.com/watch?v=RB_17SY_ngl.

³¹ Por. Warwick, dz. cyt., s. 295.

³² Por. J. Bohannon i in., *Artificial Intelligence Beats Go Champ*, w: *From AI to Protein Folding: Our Breakthrough Runners-up*, „Science”, 22.12.2016.

karzy, a LaMDA przekonuje nadzorującego go inżyniera z Google, że posiada świadomość i należy uznać go za osobę³³.

Nelson, w przywołanym wcześniej założycielskim manifestie na temat relacji technologii komputerowych i człowieka, zdecydowanie r o z r ó ż n i a ł i n t e l i g e n c j ę c z ł o w i e k a i i n t e l i g e n c j ę m a s z y n y, podkreślał rolę intencji, wolnej woli i kreatywności³⁴. Dziś wszystkie te obszary, włącznie z empatią, emocjami, inteligencją, zdolnościami komunikacyjnymi i wolną twórczością, stanowią obszar badań i eksperymentów rozwijających zarówno roboty, jak i sztuczną inteligencję. Te – dotąd postrzegane jako t y p o w o l u d z k i e – c e c h y stały się obszarem podlegającym projektowaniu (a zatem są badane, modelowane, programowane i prototypowane). Ta zmiana wiąże się z e w o l u c j ą m y ś l e n i a o s a m e j i n t e l i g e n c j i i z e z n a c z n ą z m i a n ą r o z u m i e n i a j e j i s t o t y w stosunku do pierwszych pism cybernetyki i koncepcji widzących inteligencję jako symulację procesów logicznych, na przykład gry w szachy.

Zmiana myślenia widoczna jest w obszarze projektowania robotów -androidów. Japoński konstruktor-robotyk Hiroshi Ishiguro rozwija koncepcję „senzai-kan”, charakterystyczną dla ludzi świadomości „bycia-tu”, poczucia istnienia³⁵. Roboty Ishiguro, między innymi Geminoid (wierna kopia twórcy), budzą zarazem i fascynację, i lęk³⁶. Mimo wysokiego zaawansowania programistycznego i technologicznego, nie pokonały one wciąż „doliny osobliwości” (ang. uncanny valley), zdefiniowanego przez Masahiro Mori³⁷ problemu robotyki: bliskiego podobieństwa i jednocześnie dziwnej, odczuwanej jako nieprzyjemna, obcości robota (przypominającego zombie).

Problem stanowi nie tylko dostępna technologia, ale przede wszystkim ludzka percepcja i oczekiwania poznawcze. Zwiększanie perfekcji odwzorowania ludzkiego ciała i doskonalenie sztucznej inteligencji nie przyniosły dotąd rozwiązania. Paradoksalnie, roboty wyposażone w sztuczną inteligencję, które w k o m u n i k a c j i z c z ł o w i e k i e m p r z e j a w i a j ą a u t o i r o n i ę i d z i a ł a n i a i r r a c j o n a l n e, postrzegane są przez ludzi jako bardziej ludzkie i budzą pozytywniejsze odczucia³⁸. Sugerując, że

³³ Por. B. L e m o i n e, *Is LaMDA Sentient? – an Interview*, „Medium”, 11.06.2022, <https://cajundiscordian.medium.com/is-lambda-sentient-an-interview-ea64d916d917>; M i n t z, dz. cyt.

³⁴ Por. N e l s o n, dz. cyt.

³⁵ Mówi o tym Ishiguro w wywiadzie w filmie dokumentalnym *Mechanical Love* (Finlandia, 2007, reż P. Ambo).

³⁶ Zob. H. I s h i g u r o, *Symbiotic Human-Robot Interaction Project*, <https://www.jst.go.jp/erato/ishiguro/en/index.html>.

³⁷ Zob. M. M o r i, *The Uncanny Valley* (1970), „IEEE Spectrum”, 12.06.2012, <https://spectrum.ieee.org/the-uncanny-valley>.

³⁸ Zob. M. L. T r a e g e r i n., *Vulnerable Robots Positively Shape Human Conversational Dynamics in a Human–Robot Team*, „PNAS” 117(2020) nr 12, s. 6370–6375.

także (jak ludzie) popełniają błędy, realizują pewien algorytm, dzięki któremu w ludzkim odbiorze stają się bardziej do nas podobne. Nie są niellogiczne, a jednak mogą budować takie wrażenia w interaktorach. Taka symulowana (udawana) niekonsekwencja i niedoskonałość byłyby zatem tymi elementami, które sprawiają, że technologia przestaje budzić lęk i wydaje się bliższa innym doświadczeniom komunikacyjnym (naturalnie zakorzenionym w interakcjach z ludźmi, pełnymi niekonsekwencji). Jest to początek p r o g r a m o w a n i a m a n i p u l a c j i l u d z k ą p e r c e p c j ą w kontakcie z AI.

W powiązaniu z obserwacjami różnych zespołów badawczych w zakresie kreatywności komunikacyjnej sztucznej inteligencji ukazuje to interesujący nowy aspekt „oswajania” AI. Boty wytwarzają już własne języki o charakterze tekstowym lub matematyczno-wizualnym³⁹. Ludziom wydają się one niellogiczne i niekomunikatywne, lecz prowadzą do udanych negocjacji między botami, a zatem spełniają zadanie środka komunikacji⁴⁰. Historia rozwoju dziedziny sztucznej inteligencji pełna jest zaskakujących zwrotów, jak zauważają twórcy Open AI. Trudno zatem przewidzieć, w którym kierunku będzie zmierzać rozwój tych technologii⁴¹.

Obecnie zaczynamy już rozważać sztukę robotów, sztukę AI, samoreplikację robotów, samodoskonalenie AI, uczenie się, emocje robotów i AI, świadomość, niedługo zacniemy zapewne mówić o ich duchowości. Proces ten można uznać nie tylko za zmianę w obrębie rozumienia samej inteligencji, ale też za h u m a n i z a c j ę r o b o t ó w i s z t u c z n e j i n t e l i g e n c j i. Postępuje ona wraz z powszechną akceptacją transhumanistycznej negacji wyjątkowości człowieka i transcendencji, biologizacją kultury i aktywności duchowej, umysłowej, a także aksjologii i emocjonalności człowieka w dyskursie naukowym, wraz z popularyzacją neoewolucjonizmu i memetyki⁴²,

³⁹ Zob. M. L e w i s i i n., *Deal or No Deal? End-to-End Learning for Negotiation Dialogues*, <https://s3.amazonaws.com/end-to-end-negotiator/end-to-end-negotiator.pdf>.

⁴⁰ Zob. tamże.

⁴¹ Zob. G. B r o c k m a n, I. S u t s k e v e r, O p e n A I, *Introducing Open AI*, <https://openai.com/blog/introducing-openai/>.

⁴² Neoewolucjonizm jest prądem na nowo analizującym koncepcję ewolucji, stawiającym tezę, że koncepcja K. Darwina nie jest błędna, choć bywała błędnie interpretowana. Memetyka z kolei dokonuje powiązania pomiędzy ewolucją genetyczną a transmisją kulturową i rozwojem cywilizacji oraz techniki, poszukując najmniejszej cząstki informacji, która jest przenoszona różnymi drogami we wszystkich tych procesach (por. S. J. G o u l d, *Niewczesny pogrzeb Darwina*, tłum. N. Kancewicz-Hoffman, Państwowy Instytut Wydawniczy, Warszawa 1991 [neoewolucjonistyczne podejście do ewolucji i rozwoju człowieka w świetle badań antropologicznych, paleontologicznych i archeologicznych końca XX wieku]; R. D a w k i n s, *Samolubny gen*, tłum. M. Skoneczny, Prószyński i S-ka, Warszawa 1996 [memetyczna koncepcja rozwoju organizmów na Ziemi na podstawie ewolucji genetycznej i kulturowej (memetycznej), propozycja terminu „mem”]; S. B l a c k m o r e, *Maszyna memowa*, tłum. N. Radomski, Rebis, Poznań 2002 [rozszerzenie koncepcji memetycznej o ewolucję technologiczną, propozycja wyróżnienia „temów”].

a w medycynie postrzeganiem człowieka jako bytu hormonalnego, czyli biologicznego, lecz algorytmicznego, zależnego od procesów chemicznych. Jest to kilka prądów myślowych, wśród których wyróżnia się zwrot biologiczny⁴³, transhumanizm oraz algorytmizacja poznania⁴⁴. Ich wspólnym mianownikiem jest mechanistyczne spojrzenie na człowieka oraz antyhumanizm. Powrót do biologiczności i odwrót od duchowości (a także kultury, filozofii, psychologii) można postrzegać jako *d e h u m a n i z a c j ę c z ł o w i e k a*, konieczną do „oswojenia” AI i powstania humanity+.

EWOLUCJA INTELIGENCJI: AI TO NIE KONIEC

Epoka „eksplozji inteligencji”, przewidywana w latach sześćdziesiątych przez Irvinga J. Gooda, matematyka i kryptologa, współpracownika Alana Turinga w Bletchley Park, zajmującego się teorią prawdopodobieństwa i statystyką Bayesowską, niewątpliwie albo już nadeszła, albo jest blisko. Należy oczekiwać zatem pojawienia się *s u p e r i n t e l i g e n c j i*. Zdaniem Gooda nie powinniśmy nigdy dopuścić do tego momentu, gdyż będzie to koniec inteligencji ludzkiej⁴⁵. Do tej dystopijnej myśli wracano wielokrotnie w momentach przełomu technologicznego. Uczynił to między innymi Vernor Vinge u progu lat dziewięćdziesiątych i technologii globalnej, cywilnej sieci czy Nick Bostrom w 2014 roku w kontekście rozwoju AI. Bostrom zauważył, że pionierzy sztucznej inteligencji nie podejrzewali, że ich praca mogłaby powodować jakiekolwiek zagrożenie dla ludzkości⁴⁶. Stan techniki poprzednich

⁴³ Chodzi tu o interdyscyplinarne badania, łączące różne dziedziny i tradycje. Badacze mówią tu o zwrocie biologicznym z uwagi na poszukiwania podstaw procesów i zjawisk społecznych, kulturowych czy technicznych w procesach i zjawiskach o charakterze biologicznym, takich jak ewolucja, ekosystem, prawa przyrody, systemy biologiczne. Wyróżnia się tu koncepcja konsilencji, czyli łączenia poszukiwań z różnych obszarów, które dają wspólne uogólnienia, zaproponowana przez W. Whewella w 1840 roku, a przypomniana na współczesny użytek łączenia nauk przez E.O. Wilsona w jego książce z 1998, będącej jednocześnie udaną realizacją badań w duchu konsilencji w zakresie socjobiologii (zob. E.O. W i l s o n: *Konsilencja. Jedność wiedzy*, tłum. J. Mikos, Wydawnictwo Zys i S-ka, Poznań 2002). Por. *Biological Turn. Idee biologii w humanistyce współczesnej*, red. D. Wężowicz-Ziółkowska, E. Wiczorkowska, Wydawnictwo Uniwersytetu Śląskiego, Katowice 2016.

⁴⁴ Nurt ten związany jest z problematyką big data i paradygmatem data-driven science, czyli nauki opartej na danych. Chodzi tu o przejście paradygmatów od nauki: 1) opartej na obserwacji i doświadczeniach empirycznych, 2) przez opartą na teorii i modelowaniu, 3) następnie na technikach obliczeniowych i symulowaniu procesów, aż po 4) opartą na analizie big data. Paradygmat ten obejmuje nie tylko nauki ścisłe, ale też przyrodnicze, społeczne, a nawet w coraz większym stopniu humanistyczne (zob. *The Fourth Paradigm: Data-intensive Scientific Discovery*, red. T. Hey, S. Tansley, K. Tolle, Microsoft Research, Redmont, Washington 2009).

⁴⁵ G o o d, dz. cyt., s. 31-88.

⁴⁶ B o s t r o m, dz. cyt., s. 4n.

dziesięcioleci nie pozwalał na takie rozważania poza dystopiami nurtu science-fiction. Poza tym rozwój następował stopniowo, dając stosunkowo wąskie możliwości do nadużycia nowej technologii. Musk, podobnie jak Bostrom czy Ajay Agrawal twierdzą, że obecnie sytuacja wygląda inaczej: zbliżamy się do wynalezienia ogólnej sztucznej inteligencji⁴⁷.

Ważną kwestią pozostaje jednak w y o b r a ż n i a n a u k o w c ó w r e p r e z e n t u j ą c y c h n a u k i ś c i ś l e, skupionych na liczbach, algorytmach, rozwiązaniach kolejnych wąskich problemów, którzy nie patrzą często poza horyzont tego, co aktualnie możliwe⁴⁸. Naukowcy zajmujący się sztuczną inteligencją funkcjonują w ramach określonego paradygmatu nauki. Dzięki naukom informatycznym, obok wcześniej funkcjonujących paradygmatów nauk ścisłych (empirycznego, teoretycznego i komputacyjnego), wypracowano paradygmat datalogiczny (ang. data-driven, datalogic), oparty na współpracy, usieciowieniu i analizie wielkich zbiorów danych (ang. big data)⁴⁹. Od wczesnego eksperymentowania, tworzenia modeli i symulacji naukowcy w wielu dziedzinach powszechnie przeszli do „karmienia algorytmów” określonymi zbiorami danych, co przyspiesza rozwiązywanie problemów naukowych i stawianie prognoz opartych na danych (a więc wiarygodnych). Nastąpiło to w efekcie realnego przyspieszenia rozwoju oprogramowania opartego na sieciach neuronowych. Stopniowo przemiany te doprowadziły do algorytmizacji kultury i życia społecznego, nieodwracalnego procesu wykraczającego poza obszar nauki.

Skoro określa się współczesność jako „eksplozję inteligencji” i nadejście „epoki superinteligencji”, w a r t o d o o k r e ś l i ć r o z u m i e n i e i n t e l i g e n c j i. Nie jest to bez znaczenia, przez lata bowiem znaczenie terminu uległo ewolucji. Podejście na początku rozwoju teorii informacji zakładało jako inteligentne takie zachowanie, które wiąże się z wykorzystaniem wnioskowania logicznego⁵⁰. Już teksty Turinga i postawione w nich problemy pozwalają jednak na co najmniej podwójne podejście do proce-

⁴⁷ Istnieją dwa sposoby rozumienia sztucznej inteligencji, co powoduje pewne nieporozumienia. Najczęściej mamy do czynienia ze słabą SI (ang. narrow, weak, applied AI), gdy mowa o konkretnych rodzajach oprogramowania, ale to silna SI (ang. strong, general, full AI) jest przedmiotem obaw ludzkości. Słaba SI (wąska, stosowana) jest wyspecjalizowana w wąskim zakresie obliczeniowym – to na niej skupiali się przez lata badacze. Silna SI (ogólna, pełna) to inteligencja maszynowa rozumiana szeroko: samodzielna, rozwinięta, samoświadoma. To ona może być porównywana z myśleniem ludzkim, jest pewnym horyzontem poznawczym dla naukowców i twórców science-fiction (zob. A g r a w a l i i n., *Prediction Machines*).

⁴⁸ Zob. G. B e l l, T. H e y, A. S z a l a y, *Beyond the Data Deluge*. „Science” 323(2009) nr 5919, <https://science.sciencemag.org/content/323/5919/1297>.

⁴⁹ Por. przyp. 44.

⁵⁰ Por. *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life: Plus The Secrets of Enigma*, red. B.J. Copeland, Oxford University Press, Clarendon Press, Oxford 2004, s. 431.

su myślenia: (1) w orientacji związanej z teorią silnej sztucznej inteligencji poprzez myślenie rozumie się p r o c e s y l o g i c z n e i o p e r a c j e o b l i c z e n i o w e, a celem jest powielenie działania umysłu (symulacja); (2) w orientacji związanej z teorią słabej sztucznej inteligencji poprzez myślenie rozumie się natomiast skomplikowany i wielowarstwowy proces, w którym biorą udział n i e t y l k o l o g i k a, a l e i p e r c e p c j a (m u l t i - s e n s o r y c z n o ś ć), p r z e t w a r z a n i e b o d ź c ó w i e m o c j e, zatem celem modelowania jest tworzenie symulacji, które można traktować jako podstawę do weryfikowania pewnych hipotez na temat samej natury procesów myślowych, ale nie jako pełnoprawny model umysłu. W zależności od podejścia maszynom można przypisać zdolność myślenia (strong AI⁵¹) lub nie (weak AI). Turing reprezentował pierwsze podejście, ale rozważał też drugie (tu pojawiają się słynne „truskawki Turinga” w jego „grze imitacyjnej”)⁵¹. W wykładzie w London Mathematical Society w 1947 roku matematyk zauważył jednak, że „zachowanie inteligentne” to takie, w którym udział bierze też pewna nieprzewidywalność, aleatoryczność czy randomowość⁵². A zatem: n i e t y l e l o g i k a i r u t y n a, i l e o d e j ś c i e o d n i c h – ś w i a d c z ą o i n t e l i g e n t n y m, c z y l i t w ó r c z y m p o d e j ś c i u d o d a n e g o p r o b l e m u. Badacz zwrócił też uwagę na to, że stan obserwatora i jego odczucia (czyli interpretacja przewidywalności bądź nieprzewidywalności danego zachowania interlokutora) mają znaczenie dla oceny zachowania jako inteligentnego lub nie⁵³. Turing wyraził również przekonanie, iż z czasem maszyna „będzie się sama uczyć na bazie doświadczenia”⁵⁴ – można dostrzec tu wizję uczenia maszynowego.

Za tymi konceptualizacjami podążyli kontynuatorzy badań Turinga. Minsky zwrócił uwagę na wielotorowość myślenia ludzkiego: m y ś l e n i e a b s t r a k c y j n e, k o n c e p t u a l n e, z a d a w a n i e p y t a ñ i u c z e n i e s i e⁵⁵. Zajmował się także tym, jak uczą się dzieci, by móc ekstrapolować tę wiedzę w obszar inżynierii i robotyki. Wiele uwagi poświęcił s a m o ś w i a d o m o ś c i o r a z e m o c j o m jako ważnym elementom pozwalającym „myśleć po ludzku”. Jego definicja „inteligencji” zakłada „umiejętność rozwiązywania trudnych problemów” z łatwością i efektywnością⁵⁶. Inteligencją

⁵¹ Tamże, s. 463.

⁵² Tamże, s. 375.

⁵³ Por. tamże, s. 431.

⁵⁴ Tamże, s. 429.

⁵⁵ Por. M.L. M i n s k y, *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*, Simon & Schuster, New York – London – Toronto – Sydney 2006, s. 33, 281.

⁵⁶ Por. t e n ż e, *The Society of Mind*, Simon & Schuster, New York – London – Toronto – Sydney 1985, s. 71.

wyróżnia się zatem człowiek, ale nie posiadają jej zwierzęta czy superorganizmy. Może cechować się nią jednak maszyna. Człowieka natomiast badacz-transhumanista uważa za „maszynę emocjonalną”. Zwraca uwagę na problem z definiowaniem inteligencji, którą utożsamia z niezrozumiałymi procesami mentalnymi⁵⁷.

Dzięki pewnej wieloznaczności oraz zmianie rozumienia pojęcia „inteligencji” ten obszar badawczy pozostaje interesujący, co przyczynia się do dalszych poszukiwań i aplikacji teorii (nie tylko w obszarze AI, ale też w neurologii, neurokognitywistyce i psychologii). Wielu badaczy mówi zresztą o ewolucji inteligencji *per se* (a nie: rozumienia jej znaczeń) od dawna – zwłaszcza ewolucjoniści, neoewolucjoniści i memetycy. Odkrywamy dziś nowe wątki, mające istotne znaczenie dla rozumienia prehistorii człowieka, jego relacji z otoczeniem oraz rozwoju komunikacji (badania etologów, paleogenetyków, antropologów)⁵⁸. Nie chodzi tu jednak tylko o przejście od inteligencji biologicznej do sztucznej (a w momencie wyzwolenia spod kontroli człowieka do superinteligencji), a o zrozumienie pierwszej fazy rozwoju – ewolucji inteligencji biologicznej. Kolejnym stadium jest inteligencja sztuczna i rozważane w obrębie informatyki oraz nauk technicznych znaczenia, o czym była już mowa. Trzeci etap, przez niektórych badaczy uważany za przejściowy, przez innych za docelowy, to *inteligencja hybrydyczna* (rozszerzona). To w niej, czyli w połączeniu inteligencji biologicznej – ludzkiej i sztucznej – maszynowej, Edward Feigenbaum, twórca teorii systemów ekspertowych, widzi rozwiązanie niedostatków człowieka i maszyny⁵⁹. Inteligencja rozszerzona (ang. *augmented intelligence*) może stanowić najlepsze rozwiązanie, zapewniające bezpieczeństwo interesom człowieka, a jednocześnie efektywność większą niż inteligencja czysto biologiczna⁶⁰.

OSWAJANIE SZTUCZNEJ INTELIGENCJI – OSWAJANIE KOGO?

Badacze sytuujący się w obrębie podejścia konsiliencyjnego, zwrotu biologicznego, neoewolucjonizmu, memetyki oraz data-driven science wy-

⁵⁷ Tamże.

⁵⁸ Por. M. T o m a s e l l o, *Historia naturalna ludzkiego myślenia*, tłum. B. Kucharzyk, R. Ociepa, Copernicus Center Press, Kraków 2015, s. 64.

⁵⁹ Zob. E. A. F e i g e n b a u m, hasło „Expert Systems: Principles and Practice”, w: *The Encyclopedia of Computer Science and Engineering*, 1992, <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.34.9207&rep=rep1&type=pdf>; t e n ż e, *Knowledge Engineering: The Applied Side of Artificial Intelligence*, <https://stacks.stanford.edu/file/druid:qy055zd8682/qy055zd8682.pdf>.

⁶⁰ Zob. S. G o u r l e y, *Big Data and the Rise of Augmented Intelligence*, TEDxAuckland, YouTube, 5.12.2012, https://www.youtube.com/watch?v=mKZCa_cjbfq.

suwają interesujące hipotezy i dokonują nieoczekiwanych odkryć, które stawiają w nowym świetle ewolucję człowieka, zarówno tę dotychczasową, jak i przyszłą. Inspirujące dla zrozumienia procesu osławiania sztucznej inteligencji są między innymi badania prowadzone w Duke University przez antropologów Michaela Tomasella czy Briana Hare'go nad zachowaniami zwierząt, ewolucją człowieka, rozwojem komunikacji grupowej i koncepcją samoudomowienia⁶¹. Dla procesu osławiania sztucznej inteligencji koncepcja samoudomowienia niesie interesujące konsekwencje. Tłumaczyć może nie tylko psychologiczne mechanizmy związane z lękami wobec nowych technologii (reprezentującej w tym przypadku „inny” gatunek), procesy i dyskursy humanizacji maszyn i technicyzacji człowieka, ale i zaobserwowaną niedawno tendencyjność sztucznej inteligencji. Tę ostatnią tłumaczy się „biasami” danych, którymi została ona „nakarmiona”⁶², jednak w szerszym ujęciu jest to równocześnie reprezentacja konsekwencji naszej ewolucji w kierunku samoudomowienia. Dane pozyskiwane z sieci społecznościowych czy jakichkolwiek innych źródeł (w tym wypadku tekstów kultury) ukazują właśnie ślady wypracowanych w toku ewolucji zachowań obronnych, uprzedzających potencjalne zagrożenie wspólnoty przez „obcych”. W kontekście trenowania sztucznej inteligencji istotne jest uświadomienie

⁶¹ Zakłada ona, że w toku ewolucji człowiek nie tyle ośwoił dzikie zwierzęta, ile raczej ośwoiły się one same w wyniku specyficznej autoselekcji osobników. Zwierzęta podążające za produkującymi znaczne ilości odpadków grupami ludzkimi były lepiej odżywione i rozmnażały się lepiej niż osobniki zupełnie dzikie, bardziej lękliwe lub agresywne wobec człowieka. W rezultacie sukces ewolucyjny zapewniły sobie osobniki przyzwyczajające się stopniowo do obecności człowieka, czyli coraz bardziej udomowione. Rozwijały się one we własnym kręgu, nie łącząc się z bardziej lęklivymi i agresywnymi, wzmacniając tym samym w procesie selekcji geny, które odpowiadały za przyjazne zachowania wobec człowieka, wpłynęły z czasem na wygląd zwierząt oraz wywołały trwałe zmiany behawioralne związane z polepszeniem komunikacji międzygatunkowej. Badacze znajdują liczne dowody potwierdzające koncepcję samoudomowienia. Hare zauważa, że może ona tłumaczyć także pewne zmiany w zachowaniach i wyglądzie homo sapiens, który przeszedł podobny proces, jednak w obrębie własnego gatunku (rywalizując z innymi rodzajami homo). Badacze zauważają jednak, że przyjacielskie nastawienie wobec innych osobników danego gatunku nie działa wobec wszystkich ludzi, a jedynie w pewnym obszarze wspólnoty rozpoznawanej jako własna (wspólnoty kulturowej, cywilizacyjnej, etnicznej i tym podobne). Wzrastająca przyjacielskość wobec „swoich” daje bowiem w rezultacie wzrastającą wrogość wobec „obcych”, wynikającą z potrzeby obrony wspólnoty w razie niebezpieczeństwa (Hare tym przystosowaniem ewolucyjnym tłumaczy też konflikty etniczne i wojny). Zob. B. Hare, V. Woods, *Przetwarzają najzwyklejsi. Jak ewolucja wyjaśnia istotę człowieczeństwa?*, tłum. K. Kalinowski, Copernicus Center Press, Kraków 2022. Zob. też: Tomasello, dz. cyt.

⁶² Por. A. Abid, M. Farooqi, J. Zou, *Large Language Models Associate Muslims with Violence*, „Nature Machine Intelligence” 3(2021), s. 461-463, <https://www.nature.com/articles/s42256-021-00359-2>; A. Abid, M. Farooqi, J. Zou, *Persistent Anti-Muslim Bias in Large Language Models*, arXiv:2101.05783v2 [cs.CL], January 18, 2021, <https://arxiv.org/pdf/2101.05783.pdf>.

sobie tendencyjności AI oraz takie jej przygotowanie, by w fazie aplikacji nie stymulowała zachowań dehumanizujących. W dalszej perspektywie może mieć to znaczenie dla przetrwania człowieka. Skoro AI może dehumanizować część ludzi, może dehumanizować ich jako gatunek, zyskawszy świadomość swej odrębności. Niektórzy badacze AI już donoszą, że przekraczamy ten próg⁶³.

Interesujących tropów dostarcza też memetyka, którą z koncepcją samoudomowienia łączy podejście do technologii, obie postrzegają ją bowiem jako narzędzie przyspieszające przemiany ewolucyjne. Memetyka, zwłaszcza w ujęciu Susan Blackmore⁶⁴, wychodzi z założenia, że technologia (a konkretnie jej najmniejsze elementy, czyli temy) wywołuje nowy rodzaj selekcji ewolucyjnej, analogicznej do ewolucji kulturowej (memów), a jeszcze wcześniej biologicznej (genów). Koncepcja samoudomowienia wyrasta natomiast z neoewolucjonizmu i konsiliencyjnego podejścia do nauk. Technologia, o jakiej tu mowa, to prehistoryczne narzędzia i zmiana organizacji społecznej i kultury. Badacze próbując zweryfikować hipotezę o samoudomowieniu człowieka, zwracają uwagę na fakt, że na uzyskanie przewagi nad innymi gatunkami wpływ miały pogłębiane dzięki selekcji zdolności komunikacyjne, które zapewniały lepszą, czyli zarówno dłuższą, jak i szerszą socjalizację, a ta z kolei dawała przewagę technologiczną i dostęp do wiedzy, która w większej populacji podlegała cyrkulacji oraz akumulacji i transmisji międzypokoleniowej. Obydwa te procesy przyspieszyły postęp technologiczny, stając się jego kołem zamachowym. Memetyka patrzy na ten proces z perspektywy przekazu informacji, jednak zasadniczo zgadza się z tymi konstatacjami.

Takie podejście do ewolucji inteligencji zakłada zatem pewną ciągłość między inteligencją biologiczną (człowieka) a jego wytworami (kulturą, technologią). Sztuczna inteligencja jest kolejnym wcieleniem inteligencji ludzkiej, niekoniecznie opozycją. To nie tyle metafora, co uzewnętrzniony wynik procesu przetwarzania przez homo sapiens bodźców ze środowiska, z którymi mamy do czynienia od tysiącleci. Sztuczna inteligencja oznacza możliwość eksterioryzacji tych procesów i uwolnienie ich od jednostkowej woli człowieka, jednak nie tylko jest dziełem człowieka, ale też trenowana jest na dużych zbiorach danych antropologicznych, co oznacza jej bliski związek z inteligencją biologiczną, lecz w skali całych społeczności. Warto jednak pamiętać, że choć ma ona swoje błędy, biasy i halucynacje, ich źródłem nie są kody, a antropogeniczne dane.

⁶³ Emily M. Bender z University of Washington zauważa, że antropomorfizacja sztucznej inteligencji wynika z samej metaforyki pojęć stosowanych w odniesieniu do procesu uczenia się AI (por. N. Tiku, *The Google Engineer Who Thinks the Company's AI Has Come to Life*, „The Washington Post”, 11.06.2022, <http://bit.ly/3YHwGk0>; Lemoine, dz. cyt.).

⁶⁴ Zob. S. Blackmore, *The Mem Machine*, Oxford University Press, Oxford 1999 [Wydanie polskie: taż: *Maszyna memowa*, przedm. R. Dawkins, tłum. N. Radomski, Poznań Rebis, 2002].

Lęki Nelsona czy Gooda zamienione zostały dziś w konkretne działania na rzecz rozwijania nowych technologii w sposób, który pozwoli opanować negatywne emocje wobec maszyn. Mnożą się inicjatywy działań na rzecz „dobrej AI”, postępuje też edukacja w zakresie programowania sztucznej inteligencji. Maszyny dziś to oprogramowanie z przyjaznym dla użytkownika interfejsem, na przykład użyteczne aplikacje i systemy informatyczne, roboty-zabawki, które pomagają w komunikacji przedszkolakom, osobom chorym czy starszym. W takim kontekście trudno myśleć o zagrożeniu (zapomnieliśmy już o wizjach z końca poprzedniego stulecia, takich jak filmowy Terminator), gdyż maszyny i oprogramowanie oparte na AI niosą pomoc. Pewne procesy są nieuniknione, jak rozwój cyfrowej pamięci czy współlistnienie starzejących się społeczeństw z robotami pełniącymi usługi pielęgnacyjne⁶⁵. Pokonanie lęku przed technologią staje się koniecznym etapem rozwoju społecznego.

Niezależnie od tego, że póki co nowe języki chatbotów wydają nam się zabawne i irracjonalne (na przykład „fejsbotlijski” stworzony przez chatboty Facebooka⁶⁶) – niosą one jednak pewną radykalną zmianę. Nie jesteśmy już jedyni w myśleniu kreatywnym. Następnym razem, patrząc na balet w wykonaniu robota KUKA i ludzkiego tancerza⁶⁷, dostrzec możemy zatem nie tylko piękno i współpracę, ale przede wszystkim kruchość człowieka wobec ramienia robota na co dzień składającego samochody w fabrykach. Kolejne czatujące roboty nie będą nas dziwić ani bawić, podobnie jak następne zwycięstwo sztucznej inteligencji w starożytnych (ludzkich) grach⁶⁸. Sprawniejsze przetwarzanie informacji jest ewolucyjnie bardziej efektywne, a to nie musi oznaczać ludzkiego sukcesu (w obszarze intelektualnym, duchowym czy egzystencjalnym). Koncepcja memetyczna zakłada, że celem genów, memów czy temów jest przetrwanie, niezależnie od „maszyn przetrwania” (a są nimi odpowiednio: ludzkie ciała, umysły, wytwory).

Koncepcja samoudomowienia daje natomiast nadzieję, że proces osvajania sztucznej inteligencji, podobnie jak procesy osvajania innych gatunków zwierząt i samoudomowienia homo sapiens, będzie postępował w kierunku wzmocnienia życzliwości oraz przyjaźliwości. Niesie także obserwacje związane z koniecznością opanowania agresji, lęków i tych emocji, które mogą wpłynąć na dehumanizację człowieka jako gatunku lub części ludzkich społeczności. Tak jak nauczyliśmy się komunikować ze zwierzętami i ludźmi, musimy nauczyć się komunikować z nowymi formami inteligencji, czyli oprogramowaniem AI, robotem, cyborgiem i innymi hybrydycznymi formami,

⁶⁵ Zob. *Mechanical Love*.

⁶⁶ Zob. Lewis in., dz. cyt.

⁶⁷ Zob. *A Duet of Human and Robot*, Ars Electronica, Linz 2017.

⁶⁸ Por. D. Silver in., *Mastering the Game of Go with Deep Neural Networks and Tree Search*, „Nature” 2016, nr 529, s. 484-489.

które w przyszłości mogą się pojawić. Oswajanie zachodzi dwukierunkowo, działa na obydwie strony procesu komunikacji – z m i e n i a i c z ł o w i e k a, i m a s z y n ę. Już dziś rozmowy ze sztuczną inteligencją pozwalają jej twórcom na wprowadzanie ulepszeń do kodów i ograniczeń w zakresie tendencyjności. Podobnie jak przemiany umysłu i ciała człowieka pod wpływem technologii zmieniły nasz gatunek, odmieni go także komunikacja ze sztuczną inteligencją. Być może wchodzimy właśnie nie tyle w epokę sztucznej czy hybrydycznej inteligencji, co w k o l e j n y e t a p s a m o u d o m o w i e n i a. Tym razem człowiek będzie oswajał sztuczną i n t e l i g e n c j ę, a o n a c z ł o w i e k a. Może to oznaczać postępujące zbliżenie oraz symbiozę, podobną do tej, jaka zachodzi między człowiekiem a udomowionymi gatunkami zwierząt. Beneficjentami tego procesu są wszyscy, zarówno ludzie, jak i zwierzęta. Nie ma zatem powodów, by przypuszczać, że i tym razem będzie inaczej – wystarczy, by sztuczna inteligencja była nam życzliwa.

BIBLIOGRAFIA / BIBLIOGRAPHY

- Abid, Abubakar, Maheen Farooqi, and James Zou. "Large Language Models Associate Muslims with Violence." *Nature Machine Intelligence*, no. 3 (2021): 461–63. <https://www.nature.com/articles/s42256-021-00359-2>.
- Abid, Abubakar, Maheen Farooqi, and James Zou. *Persistent Anti-Muslim Bias in Large Language Models*, arXiv:2101.05783v2 [cs.CL]. January 18, 2021. <https://arxiv.org/pdf/2101.05783.pdf>.
- Agrawal, Ajay, et al. *Prediction Machines: The Simple Economics of Artificial Intelligence*. Boston: Harvard Business School Publishing, 2018.
- Bell, Gordon, Tony Hey, and Alex Szalay. "Beyond the Data Deluge." *Science*, 323 (5919) (2009). <https://science.sciencemag.org/content/323/5919/1297>.
- Beneteau, Erin, et al. "Parenting with Alexa: Exploring the Introduction of Smart Speakers on Family Dynamics." In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. New York: ACM, April 2020: 1–13. <https://doi.org/10.1145/3313831.3376344>.
- Bielak, Tomasz, and Grzegorz Ptaszek. "Algorytmiczne doświadczanie kultury." *Kultura Współczesna*, no. 1 (104) (2019): 10–14.
- Bińczyk, Ewa. "Inżynieria klimatu a inżynieria człowieka. Dyskursy na temat środowiska w epoce antropocenu." *Ethos* 28, no. 3 (111) (2015): 153–75.
- Biological Turn. Idee biologii w humanistyce współczesnej*. Edited by Dobrosława Węzowicz-Ziółkowska and Emilia Wieczorkowska. Katowice: Wydawnictwo Uniwersytetu Śląskiego, 2016.
- Blackmore, Susan. *Maszyna memowa*. Translated by Norbert Radomski. Poznań: Rebis, 2002.

- Bohannon, John, et al. "Artificial Intelligence ups its game." In From AI to protein folding: Our Breakthrough runners-up, „Science”, 22.12.2016. <https://www.sciencemag.org/news/2016/12/ai-protein-folding-our-breakthrough-runners>.
- Bostrom, Nick. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press, 2014.
- . *Superinteligencja: Scenariusze, strategie, zagrożenia*. Translated by Dorota Konowrocka-Sawa. Gliwice: Wydawnictwo Helion, 2016.
- Brief Academic Biography of Marvin Minsky, MIT. <https://web.media.mit.edu/~minsky/minskybiog.html>.
- Brockman, Greg, Ilya Sutskever, and Open AI. *Introducing Open AI*. <https://openai.com/blog/introducing-openai/>.
- Chamorro-Premuzic, Tomas. "How ChatGPT Is Redefining Human Expertise: Or How To Be Smart When AI Is Smarter Than You." *Forbes*, January 12, 2023. <http://bit.ly/3Ef5PUm>.
- Dawkins, Richard. *Samolubny gen*. Translated by Marek Skoneczny. Warszawa: Prószyński i S-ka, 1996.
- A Duet of Human and Robot*. Ars Electronica, Linz 2017. YouTube. <https://www.youtube.com/watch?v=7moBSpAEkD4>.
- Ethos* 28, no. 3 (111) (2015): *Transhumanizm*. <https://czasopisma.kul.pl/index.php/ethos/issue/view/220>.
- Feigenbaum, Edward A. *Knowledge Engineering: The Applied Side of Artificial Intelligence* (1980). <https://stacks.stanford.edu/file/druid:qy055zd8682/qy055zd8682.pdf>.
- Filo-Sofija* 17, no. 39 (1) (2017). <http://www.filo-sofija.pl/index.php/czasopismo/issue/view/47/showToc>.
- Fukuyama, Francis. "Transhumanism: A Special Report." *Foreign Policy*, October 23, 2009. <https://foreignpolicy.com/2009/10/23/transhumanism/>.
- Garbowski, Marcin. "Transhumanizm: Geneza – założenia – krytyka." *Ethos* 28, no. 3 (111) (2015): 23–41.
- Good, Irving J. "Speculations Concerning the First Ultrainelligent Machine." *Advances in Computers* 6 (1965): 31–88. <https://www.sciencedirect.com/science/article/pii/S0065245808604180>.
- Gould, Stephen J. *Niewczesny pogrzeb Darwina*. Translated by Nina Kancewicz-Hoffman. Warszawa: Państwowy Instytut Wydawniczy, 1991.
- Gourley, Sean. *Big Data and the Rise of Augmented Intelligence*. TEDxAuckland, December 5, 2012. YouTube. https://www.youtube.com/watch?v=mKZCa_ejbfq.
- Gpt Generative Pretrained Transformer, Almira Osmanovic Thunström, and Steinn Steingrímsson. "Can GPT-3 write an academic paper on itself, with minimal human input?" HAL: Open Science Archive. <https://hal.science/hal-03701250>.
- Grudowska, Joanna, and Dominik Zieliński. "Społeczeństwo 5.0: Refleksja krytyczna." *Transformacje*, no. 2 (113) (2022): 232–55.
- Hare, Brian, and Vanessa Woods. *Przetwarzają najzyczliwsi: Jak ewolucja wyjaśnia istotę człowieczeństwa?* Translated by Kasper Kalinowski. Kraków: Copernicus Center Press, 2022.

- Hiniker, Alexis et al. "Can Conversational Agents Change the Way Children Talk to People?" In *Interaction Design and Children*, ACM, June 24, 2021: 338–49. <https://dl.acm.org/doi/10.1145/3459990.3460695>.
- Humanity+. <https://www.humanityplus.org/about>.
- H+pedia [wiki], s.v. "Transhumanist Declaration." https://hpluspedia.org/wiki/Transhumanist_Declaration.
- Infonomia. *Cyborg Life: Kevin Warwick*. YouTube. April 14, 2008. http://www.youtube.com/watch?v=RB_l7SY_ngI.
- Irsak, Anna. "W obronie (tradycyjnej) medycyny." *Ethos* 28, no. 3 (111) (2015): 176–88.
- Ishiguro, Hiroshi. Symbiotic Human-Robot Interaction Project. <https://www.jst.go.jp/erato/ishiguro/en/index.html>.
- Janas, Agnieszka. *Zastosowanie agentów konwersacyjnych w nauczaniu*. „Nierówności Społeczne a Wzrost Gospodarczy”, no. 22 (2011): 281–87.
- Kleinman, Zoe. "Apple's Siri Calls Ambulance for Baby." *BBC News*, June 7, 2016. <https://www.bbc.com/news/technology-36471180>.
- Kobayashi, Kenji, and Ming Hsu. "Common Neural Code for Reward and Information Value." *PNAS* 116, no. 26 (2019): 13061–13066. <https://doi.org/10.1073/pnas.1820145116>.
- Krzysztofek, Kazimierz. "Człowiek – społeczeństwo – technologie: Między humanizmem a transhumanizmem i posthumanizmem." *Ethos* 28, no. 3 (111) (2015): 191–213.
- Kurzweil, Raymond. *The Singularity is Near: When Humans Transcend Biology*. London: Duckworth, 2009 [wydanie polskie: *Nadchodzi osobliwość. Kiedy człowiek przekroczy granice biologii*. Translated by Eliza Chodkowska and Anna Nowosielska. Warszawa: Kurhaus Publishing, 2016].
- Lemoine, Blake. "Is LaMDA Sentient? – an Interview." *Medium*, June 11, 2022. <https://cajundiscordian.medium.com/is-lamda-sentient-an-interview-ea64d916d917>.
- Lewis, Mike, et al. *Deal or No Deal? End-to-End Learning for Negotiation Dialogues*. <https://s3.amazonaws.com/end-to-end-negotiator/end-to-end-negotiator.pdf>.
- Łaszczycza, Piotr. "Człowiek i jego maszyny: Operatorzy i protezy." *Filo-Sofija* 17, no. 39 (1) (2017): 49–64.
- Maj, Anna. *Przemiany wiedzy w cyberkulturze: Badania nad kulturą, komunikacją, wiedzą i mediami*. Katowice: Wydawnictwo Uniwersytetu Śląskiego, 2021.
- Mechanical Love*. Directed by Phie Ambo. Finland 2007.
- Microsoft Learn. *Co to są agenci sztucznej inteligencji?* December 1, 2022. <https://learn.microsoft.com/pl-pl/azure/cloud-adoption-framework/innovate/best-practices/conversational-ai>.
- Miczka, Tadeusz. "Specifics of Algorithmization in Data Culture." Paper presented at the ICSRS 2019: Conference Proceedings, International Conference on Recent Social Studies and Research, CORIS Department Sapienza University, Rome, 25-26 October 2019.

- Minsky, Marvin Lee. *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. New York, London, Toronto, and Sydney: Simon & Schuster, 2006.
- . *The Society of Mind*. New York, London, Toronto, and Sydney: Simon & Schuster, 1985.
- Mintz, Steven. “ChatGPT: Threat or Menace?” *Inside Higher Ed*, January 16, 2023. <https://bit.ly/412ObwI>.
- Mori, Masahiro. *The Uncanny Valley* (1970). Translated by Karl F. MacDorman and Norri Kageki. *IEEE Spectrum*, June 12, 2012. <https://spectrum.ieee.org/the-uncanny-valley>.
- Musk, Elon. Twitter. August 3, 2014. <https://twitter.com/elonmusk/status/495759307346952192>.
- Nelson, Ted. *Computer Lib/Dream Machines*. <http://worrydream.com/refs/Nelson-ComputerLibDreamMachines1975.pdf>.
- Olzacka, Elżbieta. “Od Homo sapiens do Homo immortalis: Idea nieśmiertelności w rosyjskich projektach filozoficznych i społeczno-politycznych.” In *Technokultura: transhumanizm i sztuka cyfrowa*. Edited by Damian Gałuszka et al. Kraków: Wydawnictwo Libron, 2016.
- Open AI. <https://openai.com/about/>.
- Rheingold, Howard. *Narzędzia ułatwiające myślenie: Historia i przyszłość metod poszerzania możliwości umysłu*. Translated by Jacek B. Szporko. Warszawa: Wydawnictwa Naukowo-Techniczne PWN-WNT, 2003.
- Silver, David, et al. “Mastering the Game of Go with Deep Neural Networks and Tree Search.” *Nature*, no. 529 (2016): 484–89.
- Stelarc. <http://stelarc.org>.
- Stork, David G. “Scientist on the Set: An Interview with Marvin Minsky.” In *HAL’s Legacy: 2001’s Computer as Dream and Reality*. Edited by David G. Stork. Foreword by Arthur C. Clarke. Cambridge and London: MIT Press, 1998. <http://mitpress.mit.edu/e-books/Hal/chap2/two1.html>.
- The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life; Plus The Secrets of Enigma*. Edited by B. Jack Copeland, Oxford: Oxford University Press, Clarendon Press, 2004.
- The Fourth Paradigm: Data-intensive Scientific Discovery*. Edited by Tony Hey, Stewart Tansley, and Kristin Tolle. Redmont and Washington: Microsoft Research, 2009.
- The Encyclopedia of Computer Science and Engineering*, s.v. “Expert Systems: Principles and Practice” (by Edward A. Feigenbaum). 1992. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.34.9207&rep=rep1&type=pdf>.
- Thunström, Almira, Osmanovic. “We Asked GPT-3 to Write an Academic Paper about Itself – Then We Tried to Get It Published.” *Scientific American*, June 30, 2022. <http://bit.ly/3IDS4YM>.
- Tiku, Nitasha. “The Google Engineer Who Thinks the Company’s AI Has Come to Life.” *The Washington Post*, June 11, 2022. <http://bit.ly/3YHwGk0>.

- Tomasello, Michael. *Historia naturalna ludzkiego myślenia*. Translated by Bartłomiej Kucharzyk and Rafał Ociepa. Kraków: Copernicus Center Press, 2015.
- Traeger, Margaret L. et al. "Vulnerable Robots Positively Shape Human Conversational Dynamics in a Human–Robot Team." *PNAS* 117, no. 12 (2020): 6370–6375. <https://www.pnas.org/content/117/12/6370>.
- Warnke, Martin. "Databases as Citadels in The Web 2.0." In *Un Like Us Reader: Social Media Monopolies and Their Alternatives*. Edited by Geert Lovink and Miriam Rasch. Amsterdam: Institute of Network Cultures, 2013.
- Warwick, Kevin. *I, Cyborg*. Urbana and Chicago: University of Illinois Press, 2004.
- Weiser, Mark, and John Seely Brown. *Designing Calm Technology*. Xerox PARC, December 21, 1996. <https://people.csail.mit.edu/rudolph/Teaching/weiser.pdf>.
- Weiser, Mark. "The Computer for the 21st Century." *Scientific American* 265, no. 3 (1991): 94–104.
- Wilson, Edward O. *Konsiliencja: Jedność wiedzy*. Translated by Jarosław Mikos. Poznań: Wydawnictwo Zysk i S-ka, 2002.
- Zacher, Lech W. "Od społeczeństwa informacyjnego do społeczeństwa wiedzy (dylematy tranzycyjne: między informacją, wiedzą i wyobraźnią)." In *Spoleczeństwo informacyjne: Wizja czy rzeczywistość?* Vol. 1. Edited by Lesław H. Haber. Kraków: Akademia Górniczo-Hutnicza, 2004.

ABSTRAKT / ABSTRACT

Anna MAJ – Ewolucja inteligencji i jej badanie. Oswajanie sztucznej inteligencji w perspektywie przemian kultury i komunikacji

DOI 10.12887/36-2023-3-143-11

Celem artykułu jest przedstawienie ewolucji inteligencji (od biologicznej, przez sztuczną, po rozszerzoną) i dotychczasowych interdyscyplinarnych badań nad sztuczną inteligencją, między innymi rozwoju sposobów rozumienia terminu „inteligencja” w interesującym obszarze. W tym celu wybrano kilka podstawowych koncepcji. Ukazano także różnorodne konteksty osławiania sztucznej inteligencji w perspektywie przemian kultury i komunikacji. Wskazano na przemiany komunikacji w obrębie relacji międzyludzkich, przestrzeni domowej i publicznej, a także popkulturowych i artystycznych wizji ciała z obszaru sztuki mediów, na które znaczny wpływ mają koncepcje transhumanizmu i cyborgizacji. Wykorzystano analizę dyskursu oraz wyniki wieloletniej obserwacji uczestniczącej prowadzonej w obszarze sztuki mediów i technologii informatycznych. Ewolucji sposobów rozumienia pojęcia „inteligencja” towarzyszy zmiana w obszarze projektowania sztucznej inteligencji i samej wyobraźni twórców. Protechnologiczny dyskurs zachęca użytkowników technologii do interakcji z agentami konwersacyjnymi AI i robotami, prowadząc do humanizacji technologii, a jednocześnie do pozytywnego wartościowania maszyny oraz postrzegania cyborgizacji jako naturalnego i dobroczynnego procesu. Badania

te mogą być rozwijane w kierunku dalszej analizy przemian dyskursu dotyczącego sztucznej inteligencji, jej rozwoju i wzrastającej komplikacji relacji człowiek–technologia, na co wskazano, przedstawiając koncepcję samoudomowienia i interpretację memetyczną ewolucji.

Słowa kluczowe: sztuczna inteligencja, cyborgizacja, robotyzacja, interfejs konwersacyjny, zmiany zachowań komunikacyjnych

Kontakt: Instytut Nauk o Kulturze, Wydział Humanistyczny, Uniwersytet Śląski, ul. Uniwersytecka 4, 40-007 Katowice

E-mail: anna.maj@us.edu.pl

Tel. 32 2009323

<https://us.edu.pl/instytut/inok/osoby/anna-maj/>

ORCID: 0000-0003-3958-267X

Anna MAJ, *The Evolution of Intelligence and Its Research: Taming Artificial Intelligence, as seen from the Perspective of Changes in Culture and Communication*

DOI 10.12887/36-2023-3-143-11

The article aims to present the evolution of intelligence (from biological, through artificial, to augmented) and the existing interdisciplinary research on artificial intelligence, including developing ways of understanding the term “intelligence” in this area. For this purpose, several basic concepts have been selected. The paper also shows various contexts of taming artificial intelligence from the perspective of changes in culture and communication. The author discusses changes in communication within interpersonal relationships, home, and public space, as well as pop culture and artistic visions of the body in the area of media art, which are significantly influenced by the concepts of transhumanism and cyborgisation. The discussion is based on discourse analysis and results of many years of participant observation conducted in media art and information technology. The evolution of the ways of understanding the term “intelligence” is accompanied by a change in artificial intelligence design and in the very imagination of its creators. The pro-technology discourse encourages technology users to interact with conversational AI agents and robots, leading to the humanisation of technology and simultaneously to the positive evaluation of the machine and the perception of cyborgisation as a natural and beneficial process. Further observation and analysis of changes in the discourse on artificial intelligence, its development, and the increasing complication of the human–technology relationship will help understand the evolution of the phenomenon in question. The presented theories, the concept of self-domestication, and the memetic interpretation of evolution provide promising theoretical tools in this field.

Keywords: artificial intelligence, cyborgisation, robotisation, conversational interface, changes in communication behavior

Contact: Institute of Cultural Studies, Faculty of Humanities, University of Silesia, ul. Uniwersytecka 4, 40-007 Katowice, Poland

E-mail: anna.maj@us.edu.pl

Phone: +48 32 2009323

<https://us.edu.pl/institut/inok/osoby/anna-maj/>

ORCID: 0000-0003-3958-267X

WIZJE PRZYSZŁOŚCI

Jakub GOMULKA

ARTIFICIAL INTELLIGENCE APPLIED TO PHILOSOPHY A Contribution to the Wittgenstein Ontology Project

In 2006, the Wittgenstein Archives at the University of Bergen initiated a project that aimed at encoding Wittgenstein's Nachlass-related information into an explicit symbolic representation conforming Semantic Web technology requirements. However, their ultimate goal is to represent the subject matter of the philosopher's thought. The main purpose of the present paper is to contribute to these efforts.

For a few centuries, the word “ontology” has been used in multiple ways: a branch of philosophy that deals with what exists; a fundamental categorisation of reality; an attempt to understand being itself; or a structure of the most general possibilities. However, in recent decades, it has gained a new meaning in the context of computer science and, more specifically, in what is called Knowledge Representation and Reasoning (KRR or KR²). KR² is an offshoot of the symbolic AI research programme (also dubbed GOFAI—“good old-fashioned AI”) that was initiated in the 1950s by a group of scholars that included Allen Newell, Herbert A. Simon, and Marvin Minsky. This current dominated the AI field until around the middle of the 1990s, but it has since then been gradually ousted by the artificial neural networks paradigm that is nowadays almost synonymous with AI research as such. The proponents of symbolic AI believed that the ultimate goal of their discipline—creating a machine that would possess the ability to think like a human being (in other words, artificial general intelligence)—can be achieved by developing declarative high-level symbolic representations of problem domains as well as formal rules of reasoning on the basis of such representations. As we know today, the initial high expectations were not met due to the problem of scaling algorithmic solutions from simple toy problems to their real-life applications. Nevertheless, the knowledge representation approach was applied in useful ways in the 1970s and 1980s in the form of expert systems and still serves many purposes in both commercial and non-commercial web services and in ongoing AI research programmes such as intelligent agent approach. One offspring of the expert system paradigm is Semantic Web technology: a set of standards for data encoding, machine reasoning, and the representation of knowledge in the Internet.

To build a useful declarative representation of a piece of knowledge, one needs a shared conceptual model of a domain they want to represent; in other words, a formal explicit specification of a conceptual structure for that domain that can be known to potential users of that representation. Such a specification is called “ontology”¹.

¹ See Thomas R. Gruber, “Translation Approach to Portable Ontology Specifications,” *Knowledge Acquisition* 5, no. 2 (1993): 199.

In a broader sense, the term can be attributed to a whole knowledge base structured by a given conceptual model. In order to separate this meaning of the word from its philosophical use, in the present paper, we shall instead use the composite term “computational ontology.” Several different types of computational ontologies can be singled out: there can be general purpose conceptual models that are relatively domain-independent and include such universal categories as space or time; structures of categories that are designed to represent a specified domain or task; common sense models that serve to capture general knowledge; and also metadata ontologies that specify the vocabulary used to describe the structure of data available online.²

In 2006, the Wittgenstein Archives at the University of Bergen (WAB), an institution whose primary task is to make the Wittgenstein *Nachlass* available for researchers all around the world, initiated a project that aimed at encoding the *Nachlass*-related information into an explicit symbolic representation conforming Semantic Web technology requirements.³ Consequently, the WAB started developing a computational ontology that currently organises information about over fifty thousand *Nachlass* remarks and fragments of Wittgenstein’s published books; provides them with a link to a specific page of Wittgenstein’s manuscript, typescript, or publication; relates them to the persons they mention; and annotates them with the dates when they were written. The knowledge base thus created is available and easily searchable at the WAB website.⁴ It can be said that the project, called “Wittgenstein Ontology,” already provides metadata that can be very useful for scholars interested in the development of Wittgenstein’s philosophy. However, the aims of the WAB team⁵ are much more ambitious: their ultimate goal is to represent the subject matter of Wittgenstein’s thought.⁶ There are several reasons why such a task is difficult to accomplish: the current

² See Diana Marcela Sánchez, José María Cervero, and Esperanza Marcos Martínez. “The Road Toward Ontologies,” in *Ontologies: A Handbook of Principles, Concepts and Applications in Information Systems*, ed. Raj Sharman, Rajiv Kishore, and Ram Ramesh (New York: Springer, 2007), 9n.

³ See Alois Pichler, *Wittgenstein Ontology* (Bergen: University of Bergen, 2013), http://wab.uib.no/wab_philospace.page.

⁴ Wittgenstein Ontology Explorer, <http://wab.uib.no/sfb/>.

⁵ In the present paper, the term “WAB team” is used in a loose sense: in fact, it means a group of people involved in Wittgenstein ontology research who co-authored papers devoted to the project. In other words, it is comprised of Alois Pichler (the longtime director of the WAB), Amélie Zöllner-Weber, Jakub Mácha, Rune Falch, Matthew Fielding, Nivedita Gangopadhyay, Andreas Opdahl, Øyvind Liland Gjesdal, and a few others. Note that this list does not include all the researchers affiliated with the WAB and includes individuals who are not and never were formally affiliated with the institution.

⁶ See Alois Pichler and Amélie Zöllner-Weber, “Sharing and Debating Wittgenstein by Using an Ontology,” *Literary and Linguistic Computing* 28, no. 4 (2013): 700–707; Jakub Mácha, Rune Falch, and Alois Pichler, “Overlapping and Competing Ontologies in Digital Humanities,” in *DH-CASE '13: Proceedings of the 1st International Workshop on Collaborative Annotations in Shared Environment; Metadata, Vocabularies and Techniques in the Digital Humanities (10 September*

technology and tools that comprise Semantic Web paradigm seem to be inadequate for the representation of knowledge in the humanities. The main purpose of the present paper is to contribute to the WAB's efforts to build a comprehensive representation of Wittgenstein's philosophical legacy. All the examples discussed in the paper come from the *Tractatus Logico-Philosophicus* translated by Pears and McGuinness.⁷

We will begin with a brief outline of the basic components of the existing technology: RDF, OWL2, and SPARQL. Particularly, we will focus on the RDF reification mechanism that can prove useful in dealing with the problem of inconsistencies and different interpretations of knowledge in the humanities. Next, we will discuss the main difficulties that researchers encounter as they attempt to represent philosophical conceptions within the current paradigm. Subsequently, we will turn to the conceptual structure proposed by the WAB team, discuss the possibility of extending this structure, and try to apply WAB ideas regarding perspectives, claims, and concepts. In the course of the latter, the representation of some actual Tractarian sentences will be discussed. Particular attention will be paid to symbolic formulas. The paper concludes with an outline of possible developments of the KR² technology and their importance for the AI programme.

AN OVERVIEW OF THE ELEMENTS OF EXISTING SEMANTIC WEB TECHNOLOGY

Semantic Web (SW) aims to represent information that has a well-defined structure and can be easily searchable for classic algorithmic systems. The information stored according to the SW paradigm allows for complicated and precise query, automated processing, and generating additional information that has not been explicitly provided. The paradigm implements the following general principles:

- (1) sharing structured data in the form of documents written in a commonly accepted notation;
- (2) representing information articulated in entities with specified properties;
- (3) sharing machine-readable formal descriptions of data semantics.

The three principles are implemented, firstly, by using IRIs—Internationalised Resource Identifiers—that are usually URLs (uniform resource web addresses), to identify basic objects called resources that constitute represented knowledge. Secondly, it does so by formatting information according to the Resource Descrip-

2013, Florence, Italy), ed. Francesca Tomasi and Fabio Vitali (New York: Association for Computing Machinery, 2013); Alois P i c h l e r et al., "Crisscross Ontology: Mapping Concept Dynamics, Competing Argument and Multiperspectival Knowledge in Philosophy," *Quaderni di "Filosofia,"* no. 2 (2021): 59–73.

⁷ See Ludwig W i t t g e n s t e i n, *Tractatus Logico-Philosophicus*, trans. David F. Pears and Brian F. McGuinness (London and New York: Routledge and Kegan Paul, 1965).

tion Framework (RDF) standard in named graphs that describe relations between entities⁸. The third way the principles are implemented is serialising named graphs in one of several and mutually exchangeable formats (Turtle, N-Triples, N-Quads, JSON-LD, RDF-XML, or RDF-JSON). Lastly, this is accomplished by using formal languages to define vocabularies used in RDF descriptions; in other words, computational ontologies that consist of hierarchies of classes and properties, restrictions, and rules of generating new information from that which is given. Currently, the most popular standard for defining ontologies is OWL2⁹ combined with RDF Schema 1.1, which is an extension of RDF that was introduced in 2014.¹⁰ The most common software tool for managing information stored in SW knowledge bases is SPARQL.¹¹

The simplest piece of information represented in an RDF graph is an RDF triple that consists of three members: a subject, a property, and an object. The first of these is a resource to which the property is attributed; the last, meanwhile, can be understood as the property's value. The role of the RDF subject can be taken by an entity: a type of a resource (with a unique IRI identifier) that can stand in numerous different triples both on the position of subject and object. The role of the RDF property must be taken by a property: another type of a resource. Entities can be subsumed to types or classes by a predefined property "type," and properties can be ordered in a hierarchy. The basic structure of the triple is presented in Diagram 1.

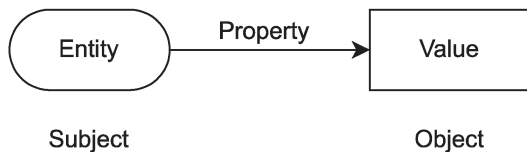


Diagram 1. The structure of the RDF triple.

There are two general types of RDF properties that differ in the kinds of values they can take. Data properties take as their values strings of Unicode characters that can have one of many predefined formats, including: date and time stamps, floating-point numeric values, or ISO language codes. Strings are not resources, and they are

⁸ See RDF Working Group. *Resource Description Framework (RDF)*. W3C. February 25, 2014. <https://www.w3.org/RDF/>.

⁹ See W3C OWL Working Group. *OWL 2 Web Ontology Language*. W3C. December 11, 2012. <https://www.w3.org/TR/2012/REC-owl2-overview-20121211/>.

¹⁰ See Dan Brickley and R.V. Guha, *RDF Schema 1.1. W3C Recommendation*, February 25, 2014, <https://www.w3.org/TR/2014/REC-rdf-schema-20140225/>.

¹¹ See Eric Prud'hommeaux and Andy Seaborne, *SPARQL Query Language for RDF, W3C Recommendation*, January 15, 2008, <https://www.w3.org/TR/rdf-sparql-query/>.

not identified by IRIs, so even if we put the same string as a value in several triples, each instance of the string will be treated as something separate.

Object properties take RDF entities as their values; in other words, object properties can be seen as binary relations between two entities. As we know, binary relations can have such features as symmetricity, reflexivity, or transitivity. All such features can be attributed to properties because the latter can take the role of the RDF subject as well. It should be noted that both object and data properties can be attributed to one entity with different values several times: an entity “Alice” can have several values of an object property “hasChild.” However, it is possible to limit a property to just one value by making it functional.

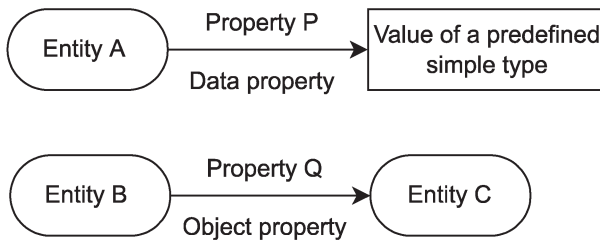


Diagram 2. Data and object properties.

We can illustrate the use of both data and object properties with examples related to Wittgenstein’s *Tractatus*. “*Tractatus Logico-Philosophicus* has 526 theses” can be represented by the entity “*Tractatus Logico-Philosophicus*,” the data property “number of theses,” and the simple integer value “526” (see Diagram 3).

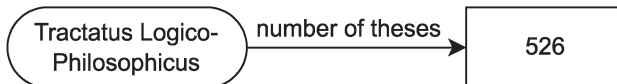


Diagram 3. An example of a data property.

In turn, the first sentence of the *Tractatus* 6.031: “The theory of classes is completely superfluous in mathematics” can be represented by the object property “redundant in” attributed to the entity “Theory of classes” with the entity “Mathematics” as its value (see Diagram 4).

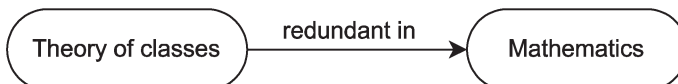


Diagram 4. An example of an object property.

The two graphs presented above can be serialised in Turtle notation as follows:

```
@prefix ex: <http://example.org/test#> .
```

```
ex:Tractatus_Logico-Philosophicus ex:number_of_theses "526" .
```

```
ex:Theory_of_classes ex:redundant_in ex:Mathematics .
```

As we can see, RDF triples usually end with a dot. The initial triple does not encode any graph; it sets a prefix to the names of all our resources used later in the code. The prefix stores a base URL address that supplements a particular resource name. In the above example, the prefix is called ‘ex’ (this name has been arbitrarily chosen) and stands for the fictitious URL address “http://example.org/test#” (it would be better if real SW knowledge bases used real URL addresses). Therefore, the actual IRI for the entity “Tractatus Logico-Philosophicus” that occurs in the second triple is “http://example.org/test#Tractatus_Logico-Philosophicus.”

Certainly, there are plenty of sentences whose structure cannot be projected into simple RDF triples without a significant loss of meaning. Fortunately, RDF specification envisages a feature that makes creating much more syntactically complex representations possible: a blank node which is not a resource and does not have an IRI. Blank nodes can stand in place of both triple’s subject and object. Therefore, they can be used to create tree-like structures that represent a more complicated syntax of a given sentence (see Diagram 5).

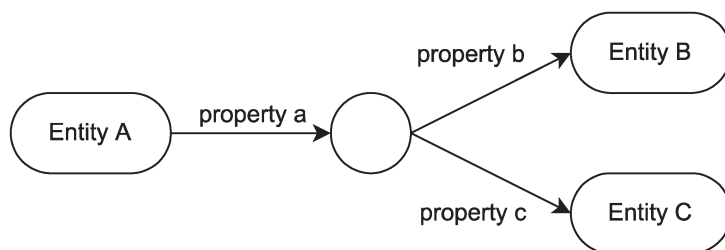


Diagram 5. A blank node structure.

For instance, if we want to represent the sentence “John travels to Vienna by train,” we can use one blank node that can represent travel (although it is possible, we do not need to name a blank node). Thus, the first RDF triple would represent the part of the information carried by the sentence; namely, that John travels. Subsequently, the blank node can have properties that specify the direction of the travel and means of transportation; therefore, the second and third RDF triples would indicate that the travel is to Vienna and that it is done by train (see Diagram 6).

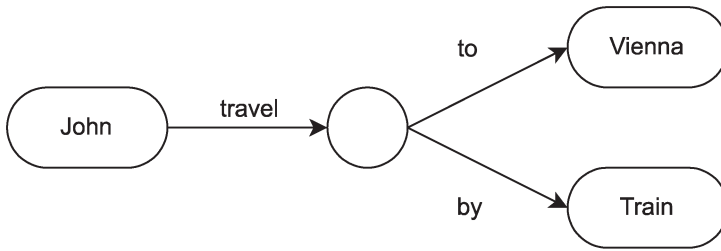
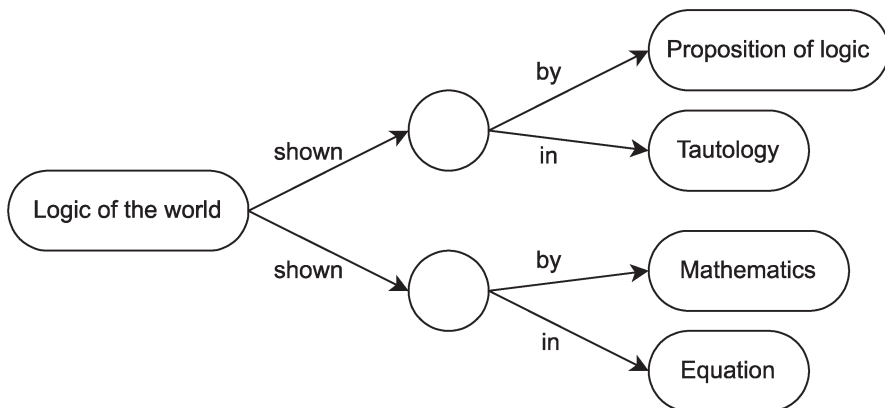


Diagram 6: A blank node example.

The Turtle symbol for a blank node is square brackets. In our example, the square brackets occur as an object in the first triple. Then, the second and third triples, whose subject is the blank node itself, appear separated by a semicolon within the brackets without their first member: by default, their subject is the blank node.

`ex:John ex:travel [ ex:to ex:Vienna; ex:by ex:Train ] .`

Equipped with this feature, we can deal with such complicated sentences as the first sentence of the *Tractatus* 6.22: “The logic of the world, which is shown in tautologies by the propositions of logic, is shown in equations by mathematics.” We can attribute the object property “shown” to the entity “Logic of the world” twice: both times the objects would be blank nodes that, in turn, have their own pairs of object properties: “by” and “in.” The two blank nodes represent the two aspects of showing; the first related to tautologies and propositional logic, the second to equations and mathematics (see Diagram 7).

Diagram 7: A named graph for the first sentence of the *Tractatus* 6.22.

The Turtle notation for the above graph is as follows (indents and new lines are only introduced to make the code clearer; they are optional elements of syntax):

```
ex:Logic_of_the_world ex:shown
[ ex:by ex:Proposition_of_logic; ex:in ex:Tautology ],
[ ex:by ex:Mathematics; ex:in ex:Equation ] .
```

As the code makes explicit, if we have two triples that share the same RDF subject and RDF property, and differ only at the RDF object, we can serialise them as one triple with both values separated with a coma.

As we will see later in this paper, from the perspective adopted in the humanities, one of the most interesting features of RDF specification is so-called reification: some RDF entities can represent RDF triples themselves. They possess special predefined RDF properties: “subject,” “predicate,” and “object.” The values of these properties can be corresponding members of a triple.¹²

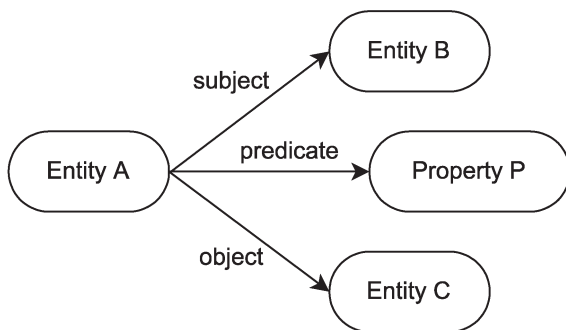


Diagram 8. The idea of RDF reification.

We can illustrate the RDF reification mechanism with another example taken from *Tractatus Logico-Philosophicus*; that is, the first sentence of thesis 6.2341: “It is the essential characteristics of mathematical method that it employs equations.” The triple that would be a subject to reification is: “mathematical method employs equations.” This fact is, as the original sentence states, the essential feature of mathematical method; consequently, the non-reified triple would consist of mathematical method as a subject, essential feature as a property, and the reified statement as an object (see Diagram 9).

¹² See Dean Allemang, Jim Hendler, and Fabien Gandon, *Semantic Web for the Working Ontologist. Effective Modeling for Linked Data, RDFS, and OWL* (New York: Association for Computing Machinery, 2020), 53–56.

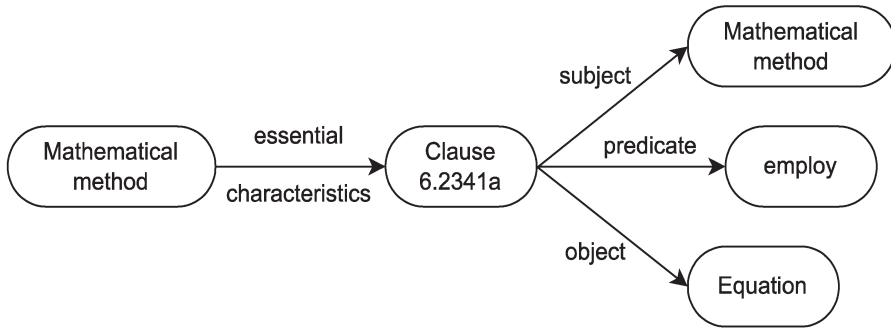


Diagram 9. An application of RDF reification

The Turtle serialisation of the above graph requires the introduction of yet another prefix because we make use of RDF predefined resources:

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
```

```
ex:Mathematical_method ex:essential_characteristics ex:Clause_6.2341a .
ex:Clause_6.2341a rdf:type rdf:Statement ;
                  rdf:subject ex:Mathematical_method ;
                  rdf:predicate ex:employ ;
                  rdf:object ex:Equation .
```

Each entity we introduce to our knowledge base can be of a certain type. In the above example, the entity “Clause_6.2341a” is a type of “Statement.” The latter is an RDF predefined class that is used for reification. However, we can easily create our own classes by declaring entities of the type “Class,” which is a predefined element of the OWL syntax. The fact that classes are also entities, and therefore resources, as well as the fact that the predefined element “Class” is itself both a resource and a class, may appear a bit confusing, but this does not lead to any real ambiguities. “Class,” together with the RDF Schema predefined property “subClassOf,” allow us to create a hierarchical categorisation; that is, a conceptual ontology, as in the following example:

```
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix wab: <http://purl.org/wittgensteinsource/ont/scholar/0.1/> .
```

```
wab:Subject rdf:type owl:Class .
wab:Source rdf:type owl:Class .
wab:Perspective rdf:type owl:Class ;
                rdfs:subClassOf wab:Subject .
wab:Primary_Source rdf:type owl:Class ;
```

```
    rdfs:subClassOf wab:Source .  
wab:WittgensteinSource rdf:type owl:Class ;  
    rdfs:subClassOf wab:Primary_Source .
```

As can be guessed from the prefixes, this is a part of a real project; that is, the Wittgenstein ontology that has been developed by the WAB.

There are numerous ways to define classes and properties in OWL, but we will not discuss this topic here. However, we should point to certain mechanisms that play their roles in the automatic production of new knowledge. Firstly, for a given property it is possible to determine the type of entities that are subjects of that property as well as the types of its values. For instance, if we define the property “is sentence of” as follows:

```
ex:isSentenceOf rdf:type owl:ObjectProperty ;  
    rdfs:domain wab:Sentence ;  
    rdfs:range wab:NachlassBemerkung .
```

then we will know that each entity to which that property is attributed is of the type “Sentence,” and each entity that is a value of that property is of the type “Nachlass Bemerkung.” Secondly, object properties can have special meta-properties that determine whether they are transitive, reflexive, irreflexive, symmetric, asymmetric, and functional and whether they are inverses of some other object properties. For instance, if we attribute the OWL predefined meta-property “TransitiveProperty” to some object property that links an entity A with an entity B and an entity B with an entity C, then we will know that it also links an entity A with an entity C. Thirdly, we can determine whether two or more distinct classes are mutually disjoint; namely, whether there is no single entity that is a member of more than one of them. So, if we ascribe a given entity to one of the group of mutually disjoint classes, we will know that the entity does not belong to any of the others.

The additional knowledge mentioned in the paragraph above can be obtained automatically as we apply a software device called a reasoner. A working reasoner extends our knowledge base with additional automatically computed RDF triples. It is worth noting that our ontologies can prove incoherent: one or more additional triples can turn out inconsistent with the existing ones. The expressivity of OWL2—much richer than that presented above—is somewhat lesser than the expressivity of the first order calculus in order to keep the reasoning process within the SW knowledge base decidable. We want to know for certain if our ontology is coherent and which additional RDF triples are computed.

SW knowledge bases would be useless if there was no tool to retrieve and manipulate the information stored in them. Just as the purpose of SQL (Structured Query Language) is the management of data held in relational databases, likewise SPARQL

(SPARQL Protocol and RDF Query Language) is a tool to manage data stored in RDF triples and structured by OWL ontologies. The most typical task for SPARQL queries is to retrieve information from a knowledge base. It can be done, just as in SQL, with the “SELECT” command followed by an indication of what should be retrieved and the ‘WHERE’ instruction followed by a clause that narrows the search placed in curly brackets. A WHERE-clause may contain an RDF graph pattern to which all results must conform. Turtle notation can be used to describe graphs, so the query can be formulated as follows:

```
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
```

```
SELECT ?p WHERE { ?p rdf:type owl:Class . }
```

A graph pattern is created when we replace at least one of the triple’s members with a variable whose name begins with the “?” sign. The query presented above will produce a list of all knowledge base entities that are classes.

We can also compose much more complicated queries by combining two or more graph patterns together and linking them with variables that do not need to occur in a SELECT-clause. Therefore, queries can be as precise as needed, and their results can be adequate. This possibility was one of the original justifications for the development of the SW technology: it seemed that the technology could provide a means to navigate effectively over the vast and chaotic collection of information that makes up the Internet. The pertinence of this line of justification faded with the rise of web search engines like Google,¹³ but it should be noted that at time of writing Google algorithms still fail to deliver adequate results in case of more complicated queries.

This brief presentation of the elements of the SW technology is far from exhaustive. Many of its elements have not been mentioned at all, and those mentioned have been discussed very concisely. Therefore, one should not treat this chapter as a comprehensive introduction to the topic but rather as a necessary context for proposals that will be discussed later.

PHILOSOPHICAL KNOWLEDGE THE HARD PROBLEM OF KNOWLEDGE REPRESENTATION?

The SW technology is best suited for representing definite knowledge. Imagine having a collection of some objects that have easily discernable features and can be

¹³ See Catherine C. Marshall, and Fran M. Shipman, “Which Semantic Web?”, in *HYPERTEXT’03: Proceedings of the Fourteenth ACM Conference on Hypertext and Hypermedia 2003*: 58f., <https://dl.acm.org/doi/proceedings/10.1145/900051>.

grouped in a clear hierarchy of classes. The knowledge about that collection can be represented without any problem; it would be a kind of catalogue. A good example of such a catalogue is the current state of the Wittgenstein ontology project: so far, the WAB team has produced over five hundred thousand RDF triples that involve over sixty-five thousand individual entities, including over fifty thousand entities for each separate remark by Wittgenstein. The remaining several thousand entities stand for persons, dates and periods of time, volumes or books, and others. They are linked together by twenty-six object properties, including “refersTo,” “hasDate,” or “hasPart,” as well as several data properties.¹⁴

Unfortunately, representing a theory, in particular, a philosophical conception is a different story. There are numerous reasons why such a task is highly problematic. The first and most important reason is the inherent multi-perspective nature of the humanities.¹⁵ The same phenomenon prompts different thinkers to present their own accounts; moreover, those accounts, in turn, prompt other scholars to present their own readings of these different accounts, bringing about new layers of divergence. Although debates are generally possible, they usually result in the formulation of yet more theories and interpretations. There is nothing wrong in this process; this is how knowledge in the humanities progresses. However, there is no clear solution to the problem of how to represent such knowledge. What are to be the entities and properties in a situation when various perspectives can offer completely different categorisations of a given phenomenon?

The second problem is the contextuality and indefiniteness of conceptions in the humanities. There is nothing strange or unnatural in the existence of their various interpretations: they are structurally open to complementing with new content and for employing in various situations that change their meanings. Therefore, knowledge in the humanities (philosophy) can never reach its definite shape and ultimate interpretation.

There is also a problem of inconsistency: some standpoints in the humanities are plainly inconsistent; that is, they contain two or more inconsistent claims. Meanwhile, others are inconsistent because it is possible to reach inconsistency through inference. Moreover, they can be inconsistent in various ways. For example, the notorious *Tractatus* 6.54 says that to understand its author is to recognise the Tractarian theses as nonsensical, as something one should throw away like a ladder. How to represent this? Is there a possibility to create a coherent computational ontology for Wittgenstein’s *Tractatus* or Hegel’s *Phenomenology of Spirit*?

Finally, the humanities are often meta-theoretical: various stances include specific categorisations of the domains about which they theorise. It is tempting to try to represent such categorisations as computational ontologies of OWL classes, but such attempts are doomed to fail because philosophical conceptual structures do not meet the formal

¹⁴ See Alois Pichler and Øyvind Liland Gjesdal, *Wittgenstein Ontology* (Bergen: University of Bergen, 2007), <http://ubdev.gitlab.io/wab-ontology/index-en.html>.

¹⁵ It should be noted that both this and other features of knowledge in the humanities (philosophy) mentioned in the main text can also be attributed, perhaps to a lesser extent, to scientific theories.

strictness requirements of first order calculus. Moreover, categorisations in the humanities are partly grounded in implicit relations between basic concepts that are informal, semantic in nature. Thus, if primitive concepts of ontologies in the humanities cannot be mapped directly onto OWL classes, what type of entities should they be? In other words, how can one build a computational ontology for a philosophical ontology?

The WAB team has long been aware of those difficulties, and they have made attempts to address them both in their theoretical papers and in the actual shape of the Wittgenstein ontology. They tested and subsequently rejected the idea of mapping philosophical categorisation onto the hierarchy of OWL classes.¹⁶ They also admitted both the possibility that various philosophical claims of the same philosopher may contradict each other, as well as the possibility that the same philosophical content may be interpreted as various conceptual structures.¹⁷ Furthermore, they have described philosophical content as dynamic, open-ended, vague, and context-dependent.¹⁸

The WAB answer to the problem of representing philosophical ontologies is the flat and limited hierarchy of classes that are directly responsible for grouping content-related entities. In turn, their solution to the other three problems is to introduce a class “Perspective” and make representations of the philosophical content dependent on it. Diagram 10 illustrates the current state of the Wittgenstein project class hierarchy. It consists of three top-level classes: “Person,” “Source,” and “Subject.” The first of them groups entities that represent persons pertinent to the content of the Wittgenstein *Nachlass* and has no child-classes. The “Source” class is the most ramified: its direct child-classes are “Primary Source” and “Secondary Source.” The former has four child-classes, including the “WittgensteinSource” class that, in turn, has nine child-classes. Those bottom-level classes are used to group such *Nachlass*-related entities as separate remarks (“*Nachlass Bemerkung*,” “Part”), manuscripts (“MS”), and typescripts (“TS”).

The third top-level class, “Subject,” has eight child-classes that constitute the bottom level of that branch.¹⁹ Four of them, “Language,” “Date,” “Place,” and “TextSub-

¹⁶ Amélie Zöllner-Weber and Alois Pichler, “Utilizing OWL for Wittgenstein’s *Tractatus*,” in *Papers of the 30th International Ludwig Wittgenstein Symposium (5–11 August 2007, Kirchberg am Wechsel)*, ed. Herbert Hrachovec, Alois Pichler and John Wang (Kirchberg am Wechsel: ALWS, 2007).

¹⁷ See Jakub Mácha, Rube Falc, and Alois Pichler, “Overlapping and Competing Ontologies in Digital Humanities,” in *DH-CASE ’13: Proceedings of the 1st International Workshop on Collaborative Annotations in Shared Environment: metadata, vocabularies and techniques in the Digital Humanities (10 September 2013, Florence, Italy)*, ed. Francesca Tomasi and Fabio Vitali (New York: Association for Computing Machinery, 2013).

¹⁸ See Pichler, Fielding, Gangopadhyay, and Opdahl, “Crisscross Ontology: Mapping concept dynamics, competing argument and multiperspectival knowledge in philosophy,” 63.

¹⁹ We have set aside the two classes the WAB team proposed in their most recent paper (see *ibidem*, 70) that are not yet present in the most recent version of the knowledge base (namely, “Debate” and “Argument”), although we are far from rejecting their usefulness. This only means that they are not pertinent to the proposals discussed in the present paper and that our primary reference is the knowledge base itself.

Genre,” are not pertinent to the philosophical content but rather to some kinds of facts regarding the documents described in the “Source” branch. The fifth, “Field,” refers to the philosophical subject matter in a very general way: it is designed to group entities that represent specific philosophical subdisciplines (currently, it includes two such entities: “Ethics” and “Philosophy of Language”). The remaining three classes that are crucial for representing philosophical knowledge are: the aforementioned “Perspective,” “Point” (or “Claim”), and “Issue” (or “Concept”). The alternative names for the last two classes indicate that the project is undergoing a transitive period regarding class naming: the most recent version of the Wittgenstein ontology OWL file,²⁰ available on December 14, 2022, still uses older versions (namely, “Point” and “Issue”), while the most recent publications use both old and new names²¹. Nowadays, meanwhile, WAB scholars are more inclined to use “Claim” and “Concept” when discussing the project.

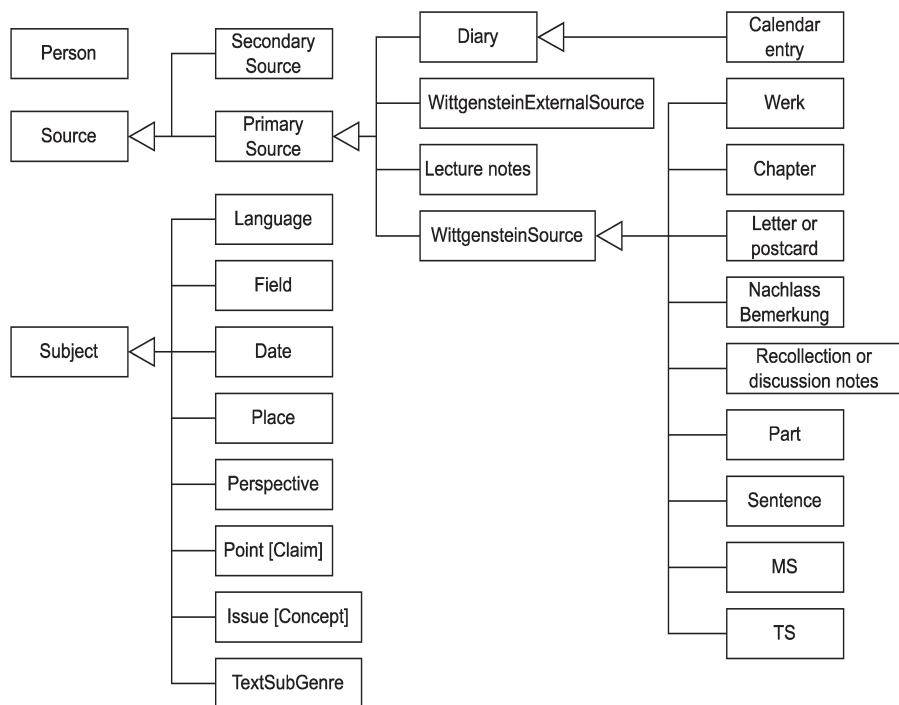


Diagram 10. The Wittgenstein ontology class hierarchy (the arrows point at the parent-classes of particular child-classes).

²⁰ The file is available for download from the web address: Wittgenstein Ontology Explorer, http://wab.uib.no/cost-a32_philospace/wittgenstein.owl (due to the large size of the file, downloading it may require a significant amount of time).

²¹ See Pichler, Fielding, Gangopadhyay, and Opdahl, “Crisscross Ontology: Mapping Concept Dynamics, Competing Argument and Multiperspectival Knowledge in Philosophy,” 59–73; Pichler and Gjesdal, *Wittgenstein Ontology*.

However, it is not quite clear how exactly the Wittgenstein ontology developers imagine relations between the exponents of the three classes; particularly, how claims and concepts will relate to perspectives. The 2021 paper suggests that node-structures for “Point” (“Claim”) and ‘Concept’ members should be developed separately: it focuses more on relations within both classes; that is, relations between different concepts and relations between different claims rather than those that link claims with concepts. It seems that the latter are simply instances of a single type of relation of occurrence of a concept in a claim.²² Diagram 11 illustrates what points (claims) and concepts are meant to represent.

It is particularly unclear what the authors mean when they write that “concepts and points shall both be allowed to develop their own sub-ontologies within the overarching general domain ontology”²³. The context suggests that this has something to do with the entities grouped in the “Perspective” class. The paper’s authors continue to write that “one will be able to merge graphs for different points which use the same concepts and are non-contradictory into larger graphs which, consequently, can represent an actual viewpoint within the specific domain,” and such larger graphs are to be perspectives: groups of non-contradictory statements by a given thinker or scholar.²⁴ However, the authors do not specify how they wish to associate an entity with a complex graph. Moreover, they make no specific remarks regarding the relation of concepts (and their interrelations) to perspectives: it seems as if a given domain of concepts is meant to be linked with a certain perspective only via a certain domain of claims.

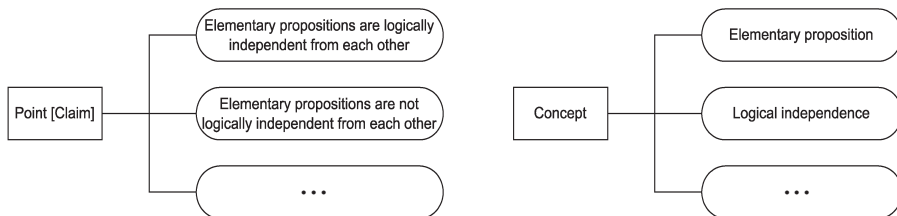


Diagram 11. The “Point” (“Claim”) and “Concept” classes (based on: P i c h l e r, F i e l d i n g, G a n g o p a d h y a y, and O p d a h l, “Crisscross Ontology: Mapping Concept Dynamics, Competing Argument and Multiperspectival Knowledge in Philosophy,” 62).

²² P i c h l e r, F i e l d i n g, G a n g o p a d h y a y, and O p d a h l, “Crisscross Ontology: Mapping Concept Dynamics, Competing Argument and Multiperspectival Knowledge in Philosophy,” 66–69.

²³ Ibidem, 69.

²⁴ Ibidem.

The 2021 WAB team paper puts forward an outline of their general task: to develop a SW knowledge base that would meet the needs of the humanities; that is, “to design novel approaches to ontology design, so that these can fully integrate humanities and philosophy contents while at the same time still retain the traditional strengths and assets of ontology work such as formal precision, cognitive economy, maximum interoperability and explanatory power, as well as permitting standard querying and inference tasks”.²⁵

In the concluding remarks, the authors admit that they have not yet figured out how exactly such a task can be accomplished.²⁶ In the next section of this paper, we will attempt to provide some detailed partial solutions. The solutions may miss the aim to retain all the traditional assets of computational ontology, but this should not be considered their flaw but rather their advantage.

PERSPECTIVES, CLAIMS, CONCEPTS, RULESETS, AND SYMBOLS

Let us commence with a couple of remarks regarding the class hierarchy proposed by the WAB. It is generally worth preserving if only for the reason that much very useful work has already been done, and the computational ontology in question is inextricably linked to the vast amount of data of which the Wittgenstein ontology project is comprised. Moreover, the approach that leaves the “Subject” branch as flat as possible seems to be right.

One of the mechanisms that will be discussed in this chapter requires a new “WittgensteinSource” child class that would group entities representing symbolic formulas used by Wittgenstein in some of his sentences. The most natural name for such a class is “Symbol,” however, the name is a bit misleading in the context of the *Tractatus*.²⁷ By “symbolic formulas,” we are referring to all strings of characters that are meant to figure in formal transformations according to a certain calculus. We can find a significant number of such strings in the *Tractatus* alone: the “[$\bar{p}$, ξ , $N(\xi)$]” expression from thesis 6, “ $(\exists x): aRx.xRb$ ” from thesis 4.1252, and also “ $0 + 1 = 1$ Def.” from thesis 6.02. The exponent of the “Symbol” class is just a node in the knowledge base; its name can be arbitrary, although it would be practical if it followed the naming convention applied in the “Source” branch and marked the volume and page of its occurrence. It should have a data property that would be similar to the proposed “shape” data property for the “Sentence” class. In fact, it can be the very same property defined for both classes. Regarding its value, it is much more convenient to use a richer notation than just a Unicode plain text, such as a very popular TEX syntax. Accordingly, the values of the “shape” property of “Symbol” entities for the formulas presented above would be:

²⁵ Ibidem, 65.

²⁶ Ibidem, 71.

²⁷ The early Wittgenstein uses two terms: “sign” and “symbol.” A sign is a set of marks of a given shape that have a syntax and a symbol is a sign with meaning. Thus, what we understand by the term “symbol” is actually the Tractarian “sign.”

" $\overline{p}$, $\overline{x}$, $N(\overline{x})$], ' $(\exists x):aRx.xRb$, ' $1 \neq 0 + 1$."

The decision to place the "Symbol" class within the "Source" branch is only provisional because it is not quite obvious that this should be its final place. On the one hand, "Symbol" members are to represent real pieces of source textual material; moreover, they should be linked to "Sentence" exponents with the aforementioned property "is part of" that, in turn, should operate within the "Source" branch only. On the other hand, as we will see, "Symbol" entities are meant to figure in subject-matter conceptual structures: symbols belong to the abstract content rather than to the textual, material form. Moreover, the "Symbol" members would play a crucial role in the aforementioned new mechanism that plainly belongs to the "Subject" domain: a mechanism to engage parts of the represented philosophical content as elements of new inference rules used by a system to generate additional information. We will discuss the details of the mechanism near the end of this section of the paper. For the moment, it should only be mentioned that it would require yet another extension of the class hierarchy: a new "Subject" child class called "Ruleset."

To summarise the remarks on the current Wittgenstein ontology class structure, we can say that it should generally remain in the present form, and its future extensions should be rather limited. The two additional classes proposed above—"Symbol" and "Ruleset"—are strictly related to the extension of the existing SW technology that is required by the new mechanism we want to introduce.

Before we delve into more detailed presentations of the possibility of binding perspectives, claims, and concepts together, let us point to two general facts. Firstly, we should differentiate two kinds of entities of which our knowledge base would consist: one would be the resources that are employed to directly represent the elements of the knowledge domain, while the other would be those that categorise, organise, and link the former. In other words, we have the subject-matter level resources and the meta-level resources. All classes in our knowledge base would work as meta-level elements, while the vast majority of entities that are not classes would be subject-matter representations. However, the application of this division to the current form of the Wittgenstein ontology project is highly problematic: both classes and properties defined in the project documentation²⁸ are in fact representations of the structure of its subject-matter; that is, the Wittgenstein *Nachlass* and Wittgenstein's publications. Nevertheless, it is useful to separate the "hard" or "general" part of the knowledge base from its numerous individual components that throng in separate classes.

The "Subject" branch of the project is much more suited for the division in question: we need to create a mechanism or structure of representation that would be applied more universally than just for Wittgenstein's thought. Therefore, the main

²⁸ See Pichler and Gjesdal, *Wittgenstein Ontology*.

“Subject” subclasses: “Perspective,” “Claim,” “Concept,” or “Ruleset” do not say anything specific about the philosophy of the author of the *Tractatus*. Similarly, some of the object and data properties that would figure in the completed knowledge base would belong to the universal structure that would allow us to grasp the variety of knowledge in the humanities. These would be the meta-level resources. However, other properties would be specific to this particular task: they would be introduced to represent aspects of some of Wittgenstein’s ideas. Therefore, they would be subject-matter level resources.

Secondly, we should be aware that each attempt at representing philosophical content by means of the SW technology would be a form of translation: it would boil down to casting the natural language of a given philosophical work onto a much less expressive semiformal system whose set of semantic elements is strictly controlled. Therefore, a knowledge base developer would have to make many interpretative decisions not only to narrow down the number of separate lexical units but also to figure out which aspect of the meaning of a given word is at play in a given phrase.

The verb “is” usually raises particularly difficult interpretative questions. If, for instance, the *Tractatus* 6 states: “The general form of a truth-function is $[p, \xi, N(\xi)]$. This is the general form of a proposition,” we are inclined to say that the “is” in the second sentence expresses the synonymy of the two complex concepts. However, the “is” in the first sentence does something else: here, the concept is juxtaposed with a complex symbol (or rather, to use Tractarian terminology, a complex sign), so we should instead read “is” as “is expressed with.” In any case, this is yet another reason why we need to include perspectives in our attempts to build representations of philosophical conceptions: each such decision can be undermined by other scholars, and they should also have a possibility to present their own visions.

Having made these preparatory remarks, we can now turn to the main issue: how the three “Subject” subclasses: “Perspective,” “Claim,” and “Concept” can be employed to the task of representing knowledge in the humanities, or, more specifically, in the philosophy of Wittgenstein. Let us begin the description from a “Perspective” class member. A perspective can be considered to be a set of beliefs, and this should be the role of an entity of that class: it should have a certain number of properties whose values would be entities representing claims. We can label this kind of property simply as “claim”; it should have its domain in “Perspective” and range in “Claim” (the class).

A value of the “claim” property (that is, a member of the “Claim” class) would be an entity with two crucial properties. We would call them “source of claim” and “structure of claim”; the value of the former would be a member of the “Sentence” class from the “Source” branch, which is how we can anchor the subject-matter in its material, textual ground. However, the most important part of this construction would be the value of the second property: it would be an entity of the built-in type “Statement.” In other words, the structure of the claim would be modelled thanks to the RDF reification syntax. Diagram 12 illustrates this strategy with a complex graph for a relatively simple example: the first sentence of the *Tractatus* 6.

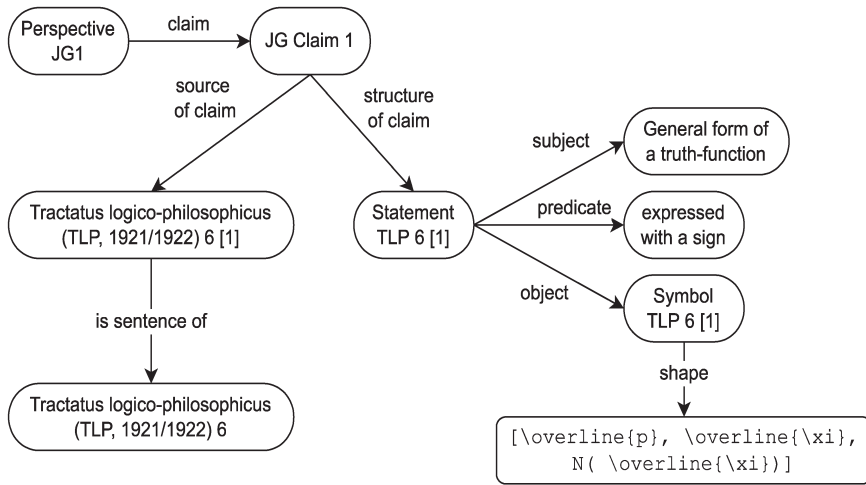


Diagram 12. The pattern of representing perspectival knowledge.

The entity in the bottom left corner of the graph is an existing member of the “Part” class that appears in the current version of the Wittgenstein ontology OWL file. All the rest is not yet present there. As can be guessed, the node “Symbol TLP 6 [1]” is meant to be a member of the “Symbol” class, and is meant to have its property “shape” filled with a TEX-style string. The element “expressed with a sign” is an object property that stands for “is” in the represented sentence as we have briefly discussed above. The definitions of the resources that figure as values of the three reification properties “subject,” “predicate,” and “object” can be given outside of the context of the perspective, although there is a reason why we should think twice about that. After all, such extra-reification definitions would repeat the old foundationalist epistemology pattern, according to which there is a level of unproblematic simple building blocks, and all meaningful disputes regarding the structures of representations require agreement about the entities occurring at that zero-level.²⁹

The first sentence of the *Tractatus* 6 is a relatively simple case of a structure that can be represented by a triple. What about more challenging examples? One of them is illustrated by the graph in Diagram 13. Here we have a possible structure of the first sentence of the *Tractatus* 6.2322 that says: “It is impossible to assert the identity of meaning of two expressions.” As we can see, the right side of the graph differs significantly from the previous example: we have a nested reification because the value of the “subject” property of the external statement is also a statement (and a blank node).

²⁹ In other words, we would fall into the trap of the Sellarsian “myth of the given” (see Wilfrid Sellars, *Empiricism and the Philosophy of Mind*, ed. Robert B. Brandom (Cambridge: Harvard University Press, 1997)).

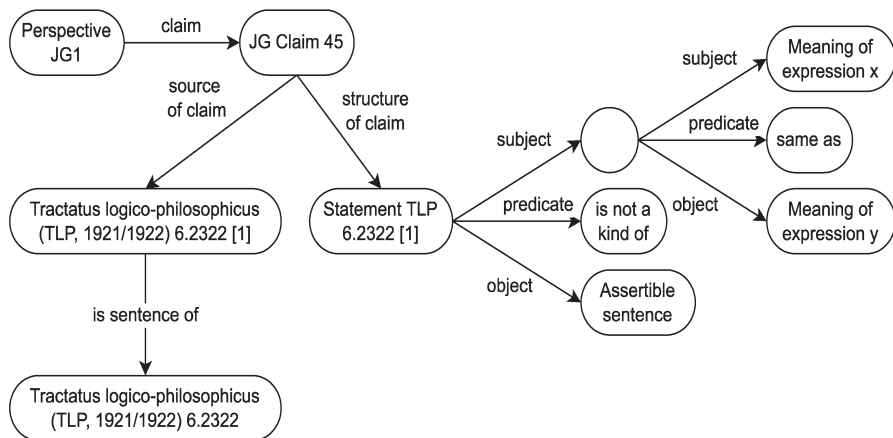


Diagram 13. Another, more complicated version of the pattern from Diagram 12.

This is only one possible variation on the theme: the starting structure from Diagram 12. We can also, for instance, multiply the “structure of claim” property and thus attribute two or more base reifications to one “Claim” entity. This should be the case for the first sentence of the *Tractatus* 6.22 that served us in the first section to illustrate the idea of a blank node (see Diagram 7).³⁰ Thus, we can make a representation of any given syntactic structure.

The application of the reification syntax has one major advantage: our knowledge base can still remain consistent despite including two (or more) claims that explicitly contradict each other because reified statements are not asserted. Moreover, in principle there is nothing that prevents us from modelling inconsistent perspectives. However, this comes at a certain price: firstly, the standard way to generate new information is blocked; secondly, queries become much more complicated as well as much less clear; thirdly, the standard SW development tools would not help in modelling such a knowledge base. We need to address the three problems now.

Let us begin with the last one. OWL files are usually not written manually in code editors: there is an application that provides a convenient graphical interface and automates some processes related to the development of a computational ontology. The application, called Protégé, was created by the Stanford Center for Biomedical Informatics Research at the Stanford University School of Medicine. It is open-source software that can be downloaded for free from its website³¹ or used as a web

³⁰ It should be evident that the graph presented in Diagram 7 cannot simply replace some part of Diagram 12. Moreover, all the blank nodes used in the reifications must be named because they usually occur in more than one reified triple.

³¹ See Protégé, <https://protege.stanford.edu>.

application. It provides a significant facilitation, especially when one is dealing with a complicated knowledge representation containing a large number of entities.

The problem is that Protégé does not support the RDF reification mechanism. Therefore, in order to be capable of developing a perspectival representation of philosophical content, one needs a different tool that would provide a graphical interface suitable for the extensive use of reifications. It should be noted that a comprehensive graphic representation of a knowledge base that includes reifications is not easy to design. There should be a kind of “surfacing” mechanism that would allow for the comprehensive inspection of information represented in the reifications related to one and the same perspective.³²

Addressing the uncertainties regarding the structure of queries over a reification-pervaded knowledge base, we can offer a similar solution: we simply need a new piece of software that would translate between complicated RDF representations and more user-friendly constructions. The software would work as an overlay on the SPARQL interface: it would take a certain form of perspective-oriented query language and produce proper SPARQL expressions that would be sent to a knowledge base constructed according to the design pattern proposed here. Let us observe that such an overlay would not affect SPARQL’s internal syntax, nor would it require any change in the query engine of a knowledge base.

The last thing we must address here is the system’s inability to generate new knowledge from knowledge represented via the reification structure. This is a rather obvious ramification of the fact that a reified triple lacks assertion. The same fact enables us to store contradictory representations in our base without making it incoherent; therefore, we should rather think of disabling the reasoning mechanism as one of the features of the proposed solution. In fact, the disabled mechanism providing inferences for non-reified chunks of information operates according to the rules of first order logic: only a few things in the humanities can be done in this way. Instead, we should think of different kinds of reasonings in the humanities and their possible implementation in the system.

Non-FOL automated reasoning is a very broad issue that can be addressed in this paper only fragmentarily. At this point, we arrive at the application of the “Ruleset” class. As has been mentioned above, it is a part of a mechanism that utilises elements of the represented content to build inference rules for generating new information. First of all, what we demonstrate here is only one of many potential ways of using that class. The way we will discuss this utilises “Symbol” entities, but much more interesting applications of the mechanism would involve concepts.

³² This would mean that the proposed application must be developed specifically for the solution outlined in Diagram 12. However, the solution is not a single-purpose mechanism; it can be applied to a broad spectrum of knowledge in the humanities.

A “Ruleset” exponent would have a number of meta-level properties. One such property would be “recursive definition.” The value of that property would be a structure made of three blank nodes. The main blank node, the direct value of the property, would be of an RDF built-in type Seq (not discussed in the section devoted to the elements of the SW technology); in other words, it would represent a sequence. In this case, the sequence would have two members given via the special RDF properties “_1” and “_2.” The other two blank nodes would be values of those properties. Their types would be left indefinite, but each of them would have two meta-level data properties: “base” and “derived.” Those properties’ values would be TEX-formatted strings. Diagram 14 illustrates the appropriate named graph.

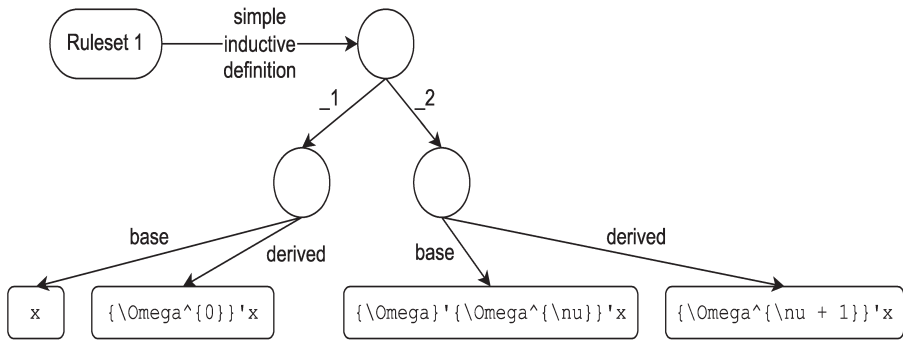


Diagram 14. A ruleset example.

As we can conjecture from the contents of the TEX-strings in the graph, the whole mechanism can be used to store recursive definitions such as the first definition in the *Tractatus* 6.02. Firstly, we provide the rule for the first member of a series (the value of the property “_1”); next, we give the rule for a general member (the value of the property “_2”). We complicate the representation of that piece of the *Tractatus* so much in order to be able to use it to automatically generate a result for any given argument. According to the first definition in 6.02, for the argument x we should get the result Ω^0x ; for Ω^1x , we should get $\Omega^{0+1}x$; for Ω^2x , we should get $\Omega^{0+1+1}x$; and so on.

Let us explain that once more, this time with the help of the Turtle notation:

```
prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
prefix ex: <http://example.com#> .
```

```
ex:Ruleset1 rdf:type ex:Ruleset .
ex:Ruleset1 ex:simple_recursive_definition [
  rdf:_1 [ ex:base "x"; ex:derived "{\Omega^{0}}'x" ];
  rdf:_2 [ ex:base "{\Omega}'{\Omega^{\nu}}'x";
           ex:derived "{\Omega^{\nu + 1}}'x" ] ] .
```

As we enter the above code in our hypothetical knowledge base, we would write the SPARQL query as follows:

```
prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
prefix ex: <http://example.com#> .

SELECT ?d WHERE { ex:Ruleset1 ex:simple_recursive_definition ?aux1 .
                  ?aux2 ex:base "{\Omega}'{\Omega}'{\Omega}'{\Omega}'x" .
                  ?aux2 ex:derived ?d .
                  ?aux1 ex:derive ?aux2 . }
```

There are four RDF triples in the WHERE-clause of the query. The first of them asks for the value of the “simple recursive definition” property that is attributed to the “Ruleset1” entity. This value is then stored in the variable “aux1.” The next triple defines an anonymous node (stored in the variable “aux2”) that has the property “base” with a value “{\Omega}'{\Omega}'{\Omega}'{\Omega}'x;” this string is our argument. The third triple links the anonymous node “aux2” with the output variable “d” by the property “derived,” so the named graph for “aux2” conforms to the named graph of the blank node that is the value of the RDF “_2” property.

The real magic happens in the last triple: here, we have the property “derive” that is not given as a property of the main blank node in the definition of “Ruleset1.” This is to be one of the magic properties in the mechanism of deriving anything with the help of “Ruleset” entities. In other words, some apparent properties must be hidden meta-keywords written in the SPARQL query engine running on the knowledge base server. As the query engine encounters one of them in the course of processing a query, it launches an appropriate script that computes a desired output value based on data that has already been retrieved. In our example, the output value would be “{\Omega}^{0+1+1+1+1}'x,” and the script that would compute recursive definitions operating on symbols’ syntax is relatively simple to write.

The main charge against solutions of that kind would certainly point to the fact that in order to implement them, we need to make modifications to the SPARQL interpreter. Unlike in the case of the RDF reification mechanism, where one only needs a new tool, here we attempt to make changes to one of the core SW technologies. This can raise concerns regarding compatibility. However, it can be noted that the proposed changes do not affect any existing SPARQL mechanisms; they only complement them with some additional capabilities. This means that the backward compatibility will be preserved.

Again, the complexity of the actual query calls for some sort of overlay software that would translate more user-friendly commands into the modified SPARQL code.

 *

It is no secret that the SW paradigm is suited to a different kind of knowledge than what one deals with in the humanities. The WAB exponents use the juxtaposition of “jigsaw puzzle” and “crisscross” knowledge: in the most SW-friendly cases, all the pieces of the represented domain have their well-defined places³³. Philosophy is very far from being such a domain: its categories are dynamic, overlapping, context-sensitive, and open to various interpretations; there is no universal agreement about virtually anything in philosophy, save maybe such simple facts as the number of the *Tractatus*’ theses. Perhaps philosophy is the least SW-friendly case.

Nevertheless, in the present paper we have discussed the prospects of a project that aims at providing an SW knowledge base that would cover the philosophy of Wittgenstein. Over the past fifteen years, a group of WAB-linked scholars has made attempts at figuring out how one can utilise RDF and OWL to produce the most adequate representation of “crisscross” knowledge. Their theoretical achievements are notable: they have come up with the idea of an ontology that allows for internal contradictions thanks to the introduction of individual perspectives. According to the WAB team, perspectives meant to embrace non-contradictory chunks of information related to the same set of concepts³⁴. The present paper attempts to flesh out the idea with the help of the RDF reification mechanism. The solution appears to be even more liberal than the theory: a single perspective can include contradictory claims due to the fact that reified RDF triples are not taken as arguments by automated reasoners; in other words, reification strips assertion from statements.

As has been noted in the preceding section of the paper, having the automated inferential mechanism disabled is not a bad thing for a knowledge base in the humanities; we should instead think of different mechanisms of reasoning that are more suitable to the subject matter. In the present paper, one such mechanism was presented: it was an idea of a reasoner that would work on the basis of a recursive definition whose specification would be dependant on the actual content of the knowledge base (in this case, the content was the first definition from the *Tractatus* 6.02).

However, the same mechanism can be applied in a very different way: we can create reasoners that utilise statistical computations and advance pattern recognition algorithms based on neural networks. We can build systems of inferences that learn to provide satisfactory answers to problems that have no definite solutions. We can apply AI deep learning algorithms. All these possibilities will be available under the

³³ See Pichler, Fielding, Gangopadhyay, and Opdahl, “Crisscross Ontology: Mapping Concept Dynamics, Competing Argument and Multiperspectival Knowledge in Philosophy,” 59–60.

³⁴ See ibidem, 69; Mácha, Falch, and Pichler, “Overlapping and Competing Ontologies in Digital Humanities.”

condition that we decide to modify the SPAQRL interpreter to let it recognise “magic” properties that can launch external scripts.

One can notice that with advanced statistical reasonings and neural networks, we are no longer dealing with SW systems in a classic sense. Firstly, it would instead be a mixed system that would combine the elements of the GOF AI with the artificial neural network paradigm. It is worth noting that such a combination is broadly recognised as the future of all AI technology.³⁵

Secondly, a reasoner would no longer provide fully consistent and precise knowledge. Moreover, it would be possible for two identical reasoners to give different answers based on a common knowledge base; furthermore, one reasoner would give different answers each time it was queried. Therefore, if we enable a reasoner to add its answers to its knowledge base every time it is queried, the knowledge base in question would develop in a unique way. One can consider such a possibility as a potential flaw. However, the idea of a software agent that meanders in a not entirely predictable way fits the philosophy much better than a rigid FOL inference system. In other words, we should consider this to be a rather promising feature.³⁶

BIBLIOGRAPHY / BIBLIOGRAFIA

- Allemang, Dean, Jim Hendler, and Fabien Gandon. *Semantic Web for the Working Ontologist. Effective Modeling for Linked Data, RDFS, and OWL*. New York: Association for Computing Machinery, 2020.
- Brickley, Dan, and R.V. Guha. *RDF Schema 1.1*. W3C. February 25, 2014. <https://www.w3.org/TR/2014/REC-rdf-schema-20140225/>.
- Gruber, Thomas R. “Translation Approach to Portable Ontology Specifications.” *Knowledge Acquisition* 5 no. 2 (1993): 199–220.
- Kautz, Henry. “The Third AI Summer: AAAI Robert S. Engelmore Memorial Lecture.” *AI Magazine*, no. 43 (2022): 105–25.

³⁵ See Henry Kautz, “The Third AI Summer: AAAI Robert S. Engelmore Memorial Lecture,” *AI Magazine*, no. 43 (2022): 118–20. Kautz notices that there can be a couple of different forms of the merge of the two approaches. According to him, the most promising approach for merging the approaches is to embed the symbolic (GOF AI) reasoning engine as a subroutine of the neural engine (see *ibidem*, 119). The approach that would result from the proposal of extending SPARQL interpreter goes in the opposite direction: the neural engine is a subroutine of the symbolic reasoner.

³⁶ I would like to thank the scholars who helped me with this paper. My thanks are due, in particular, to Alois Pichler, Matthew Fielding, Andreas Opdahl, Rune Falch, Wilhelm Krüger, and Grzegorz Nalepa; all of them were ready to answer my numerous questions regarding Semantic Web technology and the Wittgenstein ontology project. I am indebted also to Filip Mazurczak and the two anonymous reviewers for their useful comments and linguistic corrections on an earlier version of this manuscript.

- Mácha, Jakub, Rune Falch, and Alois Pichler. "Overlapping and Competing Ontologies in Digital Humanities." In *DH-CASE '13: Proceedings of the 1st International Workshop on Collaborative Annotations in Shared Environment: metadata, vocabularies and techniques in the Digital Humanities (10 September 2013, Florence, Italy)*. Edited by Francesca Tomasi and Fabio Vitali. New York: Association for Computing Machinery, 2013.
- Marshall, Catherine C., and Frank M. Shipman. "Which semantic web?" *HYPER-TEXT'03: Proceedings of the Fourteenth ACM Conference on Hypertext and Hypermedia*, (2003): 57–66.
- Pichler, Alois. *Wittgenstein Ontology*. Bergen: University of Bergen, 2013. http://wab.uib.no/wab_philospace.page.
- Pichler, Alois, and Amélie Zöllner-Weber. "Sharing and Debating Wittgenstein by Using an Ontology." *Literary and Linguistic Computing* 28, no. 4 (2013): 700–7.
- Pichler, Alois, James Matthew Fielding, Nivedita Gangopadhyay, and Andreas Lothe Opdahl. "Crisscross Ontology: Mapping Concept Dynamics, Competing Argument and Multiperspectival Knowledge in Philosophy." *Quaderni di Filosofia*, no. 2 (2021): 59–73.
- Pichler, Alois, and Øyvind Liland Gjesdal. *Wittgenstein Ontology*. Bergen: University of Bergen, 2007. <http://ubbddev.gitlab.io/wab-ontology/index-en.html>.
- Prud'hommeaux, Eric, and Andy Seaborne. *SPARQL Query Language for RDF*. W3C. January 15, 2008. <https://www.w3.org/TR/rdf-sparql-query/>.
- RDF Working Group. *Resource Description Framework (RDF)*. W3C. February 25, 2014. <https://www.w3.org/RDF/>.
- Sánchez, Diana Marcela, José María Cavero, and Esperanza Marcos Martínez. "The Road Toward Ontologies." In *Ontologies: A Handbook of Principles, Concepts and Applications in Information Systems*. Edited by Raj Sharman, Rajiv Kishore, and Ram Ramesh. New York: Springer, 2007.
- Sellars, Wilfrid. *Empiricism and the Philosophy of Mind*. Edited by Robert B. Brandom. Cambridge: Harvard University Press, 1997.
- W3C OWL Working Group. *OWL 2 Web Ontology Language*. W3C. December 11, 2012. <https://www.w3.org/TR/2012/REC-owl2-overview-20121211/>.
- Wittgenstein, Ludwig. *Tractatus Logico-Philosophicus*. Translated by David F. Pears and Brian F. McGuinness. London–New York: Routledge and Kegan Paul, 1965.
- Zöllner-Weber, Amélie, and Alois Pichler. "Utilizing OWL for Wittgenstein's Tractatus." In *Papers of the 30th International Ludwig Wittgenstein Symposium (5–11 August 2007, Kirchberg am Wechsel)*. Edited by Herbert Hrachovec, Alois Pichler, and John Wang. Kirchberg am Wechsel: ALWS, 2007.

ABSTRACT / ABSTRAKT

Jakub GOMUŁKA, AI Applied to Philosophy: A Contribution to the Wittgenstein Ontology Project

DOI 10.12887/36-2023-3-143-12

The paper focuses on the Wittgenstein Ontology Project run by the Wittgenstein Archives at the University of Bergen (WAB). The project is an attempt to apply Semantic Web (SW) technology to the task of making Wittgenstein's philosophy available as a searchable knowledge base on the Internet. The SW is one of the paradigms of Knowledge Representation and Reasoning (KR²) research and is a descendant of the symbolic AI research programme (also known as 'Good Old-Fashioned AI' or GOFAI). After a brief introduction to the SW technology, the paper discusses the main difficulties in applying it to the humanities and to the philosophy of Wittgenstein in particular. Next, it turns to the WAB team's attempts to deal with these difficulties, especially a much discussed perspectival approach that would allow for a representation of competing or contradictory claims. Finally, some original solutions are proposed: namely, the implementation of the perspectival approach with the help of the RDF reification mechanism and a new mechanism of reasoning that allows rules of inference to be generated from the content of a knowledge base. The latter solution requires a significant change in present-day technologies; that is, in the SPARQL interpreter. However, it is also a rather promising extension of the present paradigm that can potentially allow the merging of GOFAI with the artificial neural network approach.

Keywords: Ludwig Wittgenstein, Semantic Web technology, computational ontology, symbolic artificial intelligence (AI)

Contact: Faculty of Humanities, AGH University of Science and Technology, ul. Czarnowiejska 36, 30-054 Cracow, Poland

E-mail: jgomulka@agh.edu.pl

Phone: +48 12 6174365

Jakub GOMUŁKA – Sztuczna inteligencja zastosowana do filozofii. Wkład w rozwój projektu Wittgenstein Ontology

DOI 10.12887/36-2023-3-143-12

Artykuł dotyczy projektu Wittgenstein Ontology realizowanego przez Archiwum Wittgensteina przy Uniwersytecie w Bergen (WAB). Projekt ten jest próbą zastosowania technologii Sieci Semantycznej (SW) do udostępnienia filozofii Wittgensteina w internecie w postaci przeszukiwalnej bazy wiedzy. SW jest jednym z paradygmatów badawczych w ramach nurtu w informatyce zwanego Knowledge Representation & Reasoning (reprezentacja wiedzy i wnioskowanie – w skrócie KR²), będącego spadkobiercą programu badawczego

symbolicznej sztucznej inteligencji (określanego również starą dobrą sztuczną inteligencją). Po krótkim wprowadzeniu do technologii SW w artykule omówiono główne trudności w stosowaniu jej do humanistyki, a w szczególności do filozofii Wittgensteina. Następnie pokazano, w jaki sposób zespół WAB próbuje przezwyciężyć te trudności, przy czym uwagę skierowano głównie na obszernie omawiane przez członków tego zespołu podejście perspektywiczne, które pozwalałoby reprezentować rywalizujące ze sobą czy wręcz sprzeczne tezy. W głównej części artykułu zaprezentowano parę oryginalnych rozwiązań, to znaczy implementację podejścia perspektywicznego przy użyciu mechanizmu reifikacji wbudowanego w składnię RDF, a także nowy mechanizm rozumowania pozwalający tworzyć reguły inferencji na podstawie treści reprezentowanych w danej bazie wiedzy. To ostatnie rozwiązanie wymaga poważnej zmiany w aktualnie używanych technologiach, ściślej mówiąc, w interpreterze SPARQL. Jest ono jednak bardzo obiecującym rozszerzeniem obecnego paradygmatu, które pozwala na połączenie w pracach nad sztuczną inteligencją tradycyjnego podejścia symbolicznego z podejściem wykorzystującym sztuczne sieci neuronalne.

Słowa kluczowe: Ludwig Wittgenstein, technologia Sieci Semantycznej, ontologia komputacyjna, symboliczna sztuczna inteligencja

Kontakt: Katedra Technologii Informacyjnych i Mediów, Wydział Humanistyczny, Akademia Górniczo-Hutnicza im. Stanisława Staszica, ul. Czarnowiejska 36, 30-054 Kraków
E-mail: jgomulka@agh.edu.pl
Tel. 12 6174365

Leszek KORPOROWICZ

REDUKCYJNE PUŁAPKI SOCJOLOGII „MYŚLĄCYCH MASZYN”

Z pewnością zarówno zdolności „myślących maszyn”, jak i sztucznej inteligencji mają wśród swoich funkcji wiele operacji symbolicznych, konceptualnych i abstrakcyjnych, ale by sprostać systemom kompetencji symbolicznych człowieka, musiałyby być wyposażone w towarzyszące im poczucie przynależności, lojalności, rzadkie, lecz realne postawy altruizmu, odpowiedzialności. Musiałyby osiągać poziom wrażliwości etycznej, zdolnej do konstruowania, a to znaczy zarówno przekazywania, jak i transformowania swojej ponadindywidualnej tożsamości.

Paleta wyzwań, które stawia przed współczesną cywilizacją szybki rozwój badań, ale i zastosowań sztucznej inteligencji działa niczym lustro, w którym odbijają się problemy daleko wybiegające poza samą sztuczną inteligencję. Próba odpowiedzi na stawiane w związku z tym pytania może ujawnić nie tylko, jak chcemy ten typ inteligencji rozumieć, ale także to, czego w żaden sposób inteligencja tego rodzaju dokonać nie zdoła. W centrum zainteresowań nader często stawiana jest natura relacji międzyludzkich w systemie „myślących maszyn”, natura samego systemu, a w końcu jego rezultat, czyli to, co system ten czyni z ludźmi, co w nich zmienia, ku jakiemu społeczeństwu, ale przede wszystkim ku jakiemu człowiekowi i człowieczeństwu prowadzi. Nie bez znaczenia są także pytania o wymogi nowych dyscyplin, mających problemy te analizować, często myląc ich przedmiot z instrumentarium badawczym, które pozostaje w ich dyspozycji. Nic więc dziwnego, że usiłowania popadają niejednokrotnie w metaforyczność, odbiegając od tego, co „ludzkie”, i koncentrując się na przedmiocie nauk inżynierskich, środowiskowych lub przyrodniczych.

Przyczynkiem do odpowiedzi na powyższe pytania może być badanie teoretycznych i metodologicznych przesłanek, które generują nauki zajmujące się tymi zagadnieniami. Jedną z nich jest socjologia, szczególnie humanistyczna, mogąca rozjaśnić istotne sposoby oddziaływania sztucznej inteligencji i kierowanego przez nią świata maszyn na sposoby organizacji życia ludzkiego i jego kluczowe wartości. Tak rozumiana to jednak nie to samo, co przedmiotowo określana socjologia „myślących maszyn” czy socjologia ich systemu, z prostej przyczyny, iż jej przedmiotem nie byłyby zjawiska, procesy i byty społeczne, ale zaawansowana technologia. Ta ostatnia jest przedmiotem zupełnie innych nauk. Z tej różnicy wynikają ważne konsekwencje w postaci redukcji, dotyczących nie tylko współczesną socjologię, ale także inne nauki

społeczne i humanistyczne, jak na przykład tak zwaną humanistykę cyfrową. Być może wiele nieporozumień mogłoby być dość szybko wyjaśnionych, gdyby użyć bardziej adekwatnego pojęcia humanistyki wspomaganej komputerowo. Warto zauważyć, że przedmiotem tak zorientowanych dyscyplin, chcąc dosłownie trzymać się zaproponowanych dla nich nazw, nie są: człowiek, kultura i społeczeństwo, ale w znacznie większym stopniu używane przez te podmioty narzędzia. To one i możliwości ich zastosowania stają się obiektem fascynacji, a właściwy przedmiot humanistyki, która pyta o istotę i przejawy człowieczeństwa, a nie o możliwości współczesnej technologii, niezależnie od sposobów jej wykorzystania, zostaje porzucony. Użyta w tym przypadku metafora, na przykład „kulturoznawstwa cyfrowego”, które w żadnej formie nie jest poznawaniem świata cyfr, czy interesującej nas socjologii „myślących maszyn”, niepostrzeżenie zamienia nie tylko przedmiot, ale i cele tych nauk, stając się kolejnym przykładem dehumanizacji¹. Nauki humanistyczne mogą być wspierane w partykularnie i operacyjnie zdefiniowanej metodyce zbierania, porządkowania, a nawet analizy danych. Humanistyka nie powinna jednak pytać o narzędzia i ich zastosowanie, ale o człowieka, ze szczególnym uwzględnieniem atrybutu jego człowieczeństwa. Ten oczywisty, wydawałoby się, obiekt jej zainteresowań sprawia, iż koncentrujemy uwagę nie tyle na pytaniu: „czym?”, ale „kim?” jesteśmy jako istoty ludzkie, uznając funkcjonalną wagę, ale nie „przewagę” używanych przez niego narzędzi i wspomagających go urządzeń. Swoistą zamianę koncentracji poznawczej w kierunku ludzkiej przedmiotowości najlepiej wyraża pojęcie transhumanizmu, które wskazuje na programowe przesunięcie analiz od ludzkiej podmiotowości do jej uwarunkowań w świecie uzależniających technologii.

Aby przeanalizować kryjące się w tych procedurach pułapki i redukcje, warto spojrzeć na nie z perspektywy trzech, testujących możliwość jej zaistnienia, pytań. Co ważne, dotyczą one nie tylko wymiaru conceptualnego dyscypliny, ale wskazują też na dynamikę realnego procesu, bez którego jej przedmiot nie będzie miał szans zaistnienia². Pierwsze pytanie dotyczy *r e l a c j i*, które muszą istnieć między konkretnymi obiektami, czyli ludźmi, grupami, ale także organizacjami. Relacje te to różnorodne formy interakcji, a ich analiza to ważny wymiar problemu. Drugie kluczowe zagadnienie dotyczy *w i ę z i* powstających w wyniku relacji. Pojawiają się wówczas takie pytania, jak: czy mają one charakter jedynie funkcjonalny, jak opisać i zdefiniować owe funkcje, i czy zależności między maszynami i cechami tworzonego poprzez

¹ Zob. L. Korporowicz, *Źródła i typy redukcjonizmów we współczesnych analizach kultury*, „Roczniki Kulturoznawcze” 7(2016) nr 1, s. 5-30.

² Zob. *Handbook of Sociology*, red. N.J. Smelser, Sage Publication, Newbury Park, London – New Delhi 1988.

nie systemu o mniej lub bardziej abstrakcyjnym charakterze mogą tworzyć właściwe dla więzi społecznych wartości? Trzecie z kluczowych kryteriów dotyczy budowanych na bazie więzi *t o ż s a m o ś c i*, a więc czegoś więcej niż partykularnych i częściowych identyfikacji konkretnych osób w ramach odgrywanych przez nie ról. Tożsamości te w świecie społecznym nie ograniczają się przy tym do samoświadomości jednostek, ale tworzą struktury o charakterze wspólnotowym, o przenikających się zakresach odniesień, uzupełniających wyznacznikach, czasami o niewspółmiernym typie wartości³.

JAKIE RELACJE?

Czy „myślące maszyny” albo też inne obiekty współczesnego świata technologii mogą wchodzić w relacje? Oczywiście, że tak. Poziom automatyki i powiązania pracy urządzeń technicznych, a więc wymiana informacji i połączona z nią wzajemna współzależność, reaktywność i adaptacja znamionują postęp techniczny znany już w starożytności, ale intensywnie rozwijany od czasów rewolucji przemysłowej. Nie jest to jednak powód, aby relacje te uznać za rzeczywistość o charakterze społecznym. Powstają bowiem inne pytania, a mianowicie: co jest przedmiotem owych relacji, co wchodzi w obszar wymiany, a następnie współzależności. Socjologia daje pod tym względem dosyć jasne odpowiedzi, demaskując jednocześnie bezprzedmiotową wręcz charakterystykę jakiegokolwiek socjologii maszyn. To, co wchodzi bowiem w obszar relacji ludzkich, to nie tylko impulsy, bodźce i sygnały, ale nadawane im znaczenia, czasami definiowane polisemiotycznie o dużej zmienności sytuacyjnej i kontekstualnej, następnie wielopoziomowe struktury wartości powiązane z właściwymi dla nich arsenalami emocji, przeżyć, a w końcu doświadczeń transgresyjnych. Jak widać relacje społeczne to nie tylko przekazanie impulsów, ale też wymiana znaczeń i rodzaj kultury intelektualnej. Czy rzeczywistość „myślących maszyn” będzie w stanie wyjść poza sferę rozumowania, procedowania, kalkulacji, a w najbardziej rozwiniętej postaci dokonywania racjonalnej optymalizacji swoich decyzji? Jakiego typu jest to racjonalność? Czy bierze ona pod uwagę rzeczywistość afektywnych atrybutów powiązanych z nią wartości oraz kardynalną różnicę pojęcia zachowań i działań?

Perspektywa socjologiczna stawia kolejny wymóg relacjom społecznym, którym jest zdolność rozwoju działań komunikacyjnych⁴. Warto przy tym za-

³ Zob. T. P a l e c z n y, *Relacje międzykulturowe w dobie kryzysu ideologii wielokulturowości*, Księgarnia Akademicka, Kraków 2017; A. T r á b k a, *Tożsamość rekonstruowana. Znaczenie migracji w biografiiach Third Culture Kids*, Wydawnictwo Naukowe Scholar, Warszawa 2014.

⁴ Zob. J. H a b e r m a s, *Teoria działania komunikacyjnego*, t. 1, *Racjonalność działania a racjonalność społeczna*, tłum. A.M. Kaniowski, Państwowe Wydawnictwo Naukowe, Warszawa 1999.

uważyć, że komunikowanie w relacjach społecznych nie jest jedynie procesem przekazywania i transmisji informacji płynących od nadawcy do odbiorcy, co niestety jest powszechnie powielanym nieporozumieniem⁵. Jest ono działaniem interakcyjnym, które opisuje łaciński termin „communicare”, a więc aktywnością nastawioną na współudział, dzielenie się i wzajemność⁶. Koniecznie należy jednak zauważyć, że wymienione już cechy komunikowania, a więc wzajemność, wymiana i współzależność, wynikać muszą z atrybutów je poprzedzających. Są nimi: ludzka podmiotowość, intencjonalność i – oparta na świecie wartości – sprawczość.

Podmiotowość człowieka gwarantuje, że nie jest on jedynie przedmiotem sterowania i ma szansę na partnerski udział w interakcji, nawet jeśli nie będzie to relacja symetryczna. Należy więc zadać pytanie, czy świat sztucznej inteligencji jako całość relacji swoich komponentów (co obrazować może używana w tym tekście metafora świata „myślących maszyn”) posiada właściwą mu podmiotowość i wspomnianą już intencjonalną sprawczość. Być może odpowiedź na to pytanie spotka się z próbą takiej jej interpretacji, która wybroni spreparowany do jej potrzeb model rzeczywistości. Niewykluczone, że uzna ów świat i jego instrumenty za samosprawcze, a nawet samoświadome. Wybiegając nawet w tak pomyślaną przyszłość, gdy skala umiejętności i możliwości myślących maszyn przerośnie nasze wyobrażenia, interpretacja ta zderzyć się musi z fundamentalnym sensem pojęcia podmiotowości, który nie daje się zredukować do zwykłej niezależności, a nawet intencjonalnej sprawczości. Sens ten zbudowany jest bowiem na trzech innych określeniach, które trudno oderwać od człowieka i jego człowieczeństwa. Pojęciami tymi są: godność, wolność i odpowiedzialność, wymienione w niezamienialnej sekwencji jako logicznej i aksjologicznej ramie ich integracji. Kwestia ta jednoznacznie rozstrzyga problem rzeczywistości społecznej maszyn i sztucznej inteligencji poprzez negatywną odpowiedź na pytanie o godność maszyny, a co za tym idzie – jej wolność i odpowiedzialność. Faktu tego nie zmieni nawet najbardziej intensywna obecność cywilizacji technicznej w sferze kultury i wieloraki jej wpływ na świat ludzkich działań, przeżyć i emocji.

⁵ Przykład takiego typu redukcji znajdujemy w pracy Walerego Pisarka *Wstęp do nauki o komunikowaniu* (zob. Walery P i s a r e k, *Wstęp do nauki o komunikowaniu*, Wydawnictwa Akademickie i Profesjonalne, Warszawa 2008), w której autor stwierdza: „Komunikacja rozumiana jako przekazywanie informacji to tajemnica samego życia. Komunikacja rozumiana jako przekazywanie informacji za pomocą znaków to tajemnica naszego człowieczeństwa” (tamże, s. 15) i dalej: „Przekazywanie treści psychicznej, i to zarówno intelektualnej, jak i emocjonalnej, a więc tego, co się myśli, lub tego, co się czuje, przez osobnika (lub osobników) A osobnikowi B nazywamy komunikowaniem” (tamże, s. 17).

⁶ Zob. L. K o r p o r o w i c z, *Komunikacja symboliczna w rozwoju tożsamości kulturowej*, w: *Tożsamość polska w odmiennych kontekstach*, red. L. Dyczewski, D. Wadowski, Wydawnictwo KUL, Lublin 2009

Podobnie rzecz ma się z innym aspektem działań komunikacyjnych, którym jest warunkująca wymianę intencjonalność, a więc świadome zorientowanie na realizację akceptowanego i obdarzonego swoją racją celu. Gdyby pominąć ten właśnie atrybut intencjonalności, komunikowanie mogłoby być efektem przypadkowego zbiegu okoliczności wymuszających go czynników i być albo jednokierunkową percepcją, albo ekspresją treści dalekiej od wartości, które niesie ze sobą *communicare*. Być może istnieją maszyny i systemy, którym dałoby się przypisać „intencjonalność” ze względu na zaprogramowanie na realizację wyznaczonych im celów. Rodzi się jednak istotne pytanie, czy uruchomienie automatyki konkretnego programu ma coś wspólnego z intencjonalnością. Intencjonalność, podobnie jak typologia relacji, musi posiadać komponenty – znaczeniowy i emocjonalny oraz bazować na rozpoznanym zestawie wartości jako elemencie środowiska wewnętrznego uczestników interakcji komunikacyjnych. Trudno powiedzieć, że nakręcona w zegarku sprężyna porusza jego wskazówki intencjonalnie. Ku temu niezbędne jest świadome rozpoznawanie celu i sensu działania.

Kolejny atrybut komunikowania – współzależność wynika z możliwości sprawczych człowieka i grupy. Jeśli bowiem którykolwiek z uczestników relacji komunikacyjnych przyjmuje rolę zupełnie zależną od innych i staje się przedmiotem całkowitego sterowania, traci nie tylko swoją podmiotowość, ale także faktyczną możliwość współdziałania i wchodzenia w świadomie rozpoznane oraz wybierane interakcje. W odróżnieniu od prostego mechanizmu oddziaływania na siebie bodźców i zachowań atrybut sprawstwa posiada w przypadku ludzi komponent wolicjonalny. Stanowi on swoiste paliwo, które wspiera czynniki motywacyjne wynikające zarówno z osobowości, jak i doświadczeń całego dotychczasowego życia, dziedzictwa kulturowego i logotwórczych dynamizmów kultury konkretnych ludzi oraz ich wspólnot. Jeśli socjologia „myślących maszyn” miałaby wspomagać relacje, w których atrybuty te nie byłyby istotne, istnieje obawa, że ludzie mogliby być im do tego zupełnie niepotrzebni. Wolicjonalność i sprawstwo nie są bowiem strategią optymalnej realizacji celów, którą posiadać może sztuczna inteligencja, chyba że wyposaży się ją w prawo do popełniania błędów, emocjonalnego zagubienia i działania na własną szkodę. Cechy te stanowią bowiem o ważnym i wbrew pozorom rozwojowym składniku człowieczeństwa.

JAKIE WIĘZI?

Więzi powstają jako rezultat relacji. Od typu relacji zależy rodzaj więzi. W perspektywie niniejszego tekstu pojawia się ważne pytanie: czy możliwa jest społeczna więź maszyn? Z pewnością nieco bardziej skomplikowane za-

gadnienie odnosi się do możliwości społecznej więzi obiektów i wytworów sztucznej inteligencji. Perspektywa socjologiczna przychodzi tu z pomocą, wskazując na odmienność czynników budujących różne typy bytów społecznych, jak na przykład populacja, zbiorowość, grupa czy wspólnota.

W przypadku populacji nie istnieje w istocie żaden czynnik określający więzi i zależności między jego elementami. Sama populacja może być wyodrębniona na podstawie dowolnej kategorii typu formalnego, czyli: wiek, płeć, wykształcenie, miejsce zamieszkania, posiadany majątek, liczba dzieci i tak dalej. Obiektem socjologii „myślących maszyn” mogłyby być populacje – kolekcje określonego ich typu, które nie wymagają pytań o wiążące je relacje ani tym bardziej o rodzaj łączącej je więzi. Jedynym społecznym aspektem ich selekcji mogą być cechy preferowanych sposobów użycia ze względu na kompetencje lub upodobania konkretnych osób lub grup, ale wówczas zainteresowanie badawcze przenosi się z maszyn na ich użytkowników, którzy stają się właściwym przedmiotem analiz.

Istotną zmianę w sposobie wyodrębnienia typu więzi społecznych przynoszą cechy realnie istniejącej zbiorowości. Tworzące ją relacje nie wychodzą jednak poza zależności o charakterze funkcjonalnym. Z tego powodu więzi między jej uczestnikami odpowiadają doraźnym, partykularnym potrzebom i ograniczają się do najprostszych wyznaczników jej spójności, jak świadomość zamieszkiwania na tym samym terenie, lojalność wobec obowiązujących na nim praw, ale także ponadterytorialne cechy przynależności do wspólnych organizacji czy korzystania z analogicznego typu usług. O ile można wyobrazić sobie analogiczne, funkcjonalne powiązanie „myślących maszyn”, które stanowić mogą komponenty dopełniających się systemów, o tyle kluczowym problemem w określeniu ich więzi pozostaje kwestia świadomości tego typu przynależności, wzajemnej adaptacji konkretnych zachowań oraz postrzegania aspektowych choćby identyfikacji. Zależność funkcjonalną reakcji trudno jednak uznać za rozwinięty typ więzi społecznych. Redukcja przedmiotu socjologii do tego typu korelacji pozbawia ją racji istnienia. Pojawia się ona dopiero w przypadku, gdy zbiorowość przeobraża się w grupę, ale demaskuje to nieadekwatność wszelkich metafor spełniających rolę iluzji w odniesieniu do społecznej więzi świata maszyn, a tym samym wyklucza jakąkolwiek ich socjologię.

W odróżnieniu od prostej zależności funkcjonalnej grupa generuje symboliczny rezultat relacji między jej członkami. Uniwersum symboliczne grupy buduje więzi pozwalające wytwarzać i kumulować jej kapitał symboliczny⁷. Korzysta on umiejętności przekraczania uwarunkowań sytuacyjnych, a więc

⁷ Zob. *Culture and Society: Contemporary Debates*, red. J.C. Alexander, S. Seidman, Cambridge University Press, Cambridge 1990.

także temporalnych czasu teraźniejszego⁸. Odwołuje się do przeszłości i przyszłości, umożliwia wykorzystanie potencjału dziedzictwa kulturowego. Co więcej, pozwala na tworzenie więzi o charakterze abstrakcyjnym, które odwołują się do konceptualnych reprezentacji grupy, i w końcu skłania do działań transgresyjnych umożliwiających ich modyfikację⁹. Z pewnością zarówno zdolności „myślących maszyn”, jak i sztucznej inteligencji mają wśród swoich funkcji wiele operacji symbolicznych, konceptualnych i abstrakcyjnych, ale by sprostać systemom kompetencji symbolicznych człowieka, musiałyby być wyposażone dodatkowo w towarzyszące im poczucie przynależności, lojalności, rzadkie, lecz realne postawy altruizmu, odpowiedzialności. Ściśle mówiąc, musiałyby osiągać poziom wrażliwości etycznej, zdolnej do konstruowania, a to znaczy zarówno przekazywania, jak i transformowania swojej ponadindywidualnej tożsamości. Wymagają one czegoś znacznie więcej niż tylko myślenia i inteligencji. Rzeczywistość społeczna, która ograniczałaby się jedynie do tego rodzaju kompetencji, nie może stać się w pełni społeczna, a co za tym idzie – nie jest możliwa jej pełnowymiarowa socjologia. Redukcje te obrazują wagę i potrzebę identyfikowania granic tej quasi-społecznej rzeczywistości, która nader często obiecuje coś, czego nie posiada, zapożyczając wiele pojęć, metafor i wyobrażeń o charakterze antropologicznym, uczestnicząc w wyścigu, w którym nie ma już ludzi. Widać to najbardziej, gdy zapytamy o więzi charakteryzujące specyficznie dojrzały typ ludzkiej grupy, którymi są wszystkie odmiany wspólnot.

Wspólnota traktuje tworzące ją więzi nie tylko jako elementy ponadindywidualnej spójności, trwania i budowania jej zintegrowanej formy istnienia. Więzy te stają się wartościami, które poza komponentem afektywnym tworzą poczucie wspomnianej już wcześniej godności jej uczestników, a nawet pozwalają mówić o godności wspólnotowej, na przykład o godności rodziny, rodu, plemienia, grupy etnicznej czy narodu. Obecne na tym poziomie wartości mogą mieć wysoce pozainstrumentalny charakter, tak jak posiadają go wartości autoteliczne. Pozwalają one traktować konkretne elementy rzeczywistości, w tym więzi społeczne, jako cele same w sobie, niewymienialne na inne, a więc niemające swoich zamienników, niebędące przedmiotem wymiany i relatywizacji. Są nimi dla przykładu przywołane godność, honor i wolność. Ze względu na swoją rolę są nieredukowalne, bo nie można dla nich znaleźć analogii, szczególnie w świecie wszystkiego, co jest w jakiś sposób sztuczne,

⁸ Zob. E. C a s s i r e r, *Filozofia form symbolicznych*, tłum. A. Karalus, P. Parszutowicz, Wydawnictwo Marek Derewiecki, Kęty 2018.

⁹ Zob. Z. W. D u d e k, A. P a n k a l l a, *Psychologia kultury. Doświadczenia graniczne i transkulturowe*, Wydawnictwo Psychologii Kultury Eneteia, Warszawa 2008; J. K o z i e l e c k i, *Koncepcja transgresyjna człowieka*, Państwowe Wydawnictwo Naukowe, Warszawa 1987.

co naśladuje, a nawet wyręcza czy wręcz usprawia, ale tylko niektóre czynności człowieka.

Problem więzi społecznych w zasadniczy sposób demaskuje progresywną koncepcję jakiegokolwiek, nawet naturalnej i w pełni ludzkiej inteligencji, która w arsenale swoich wartości nie uwzględnia czegokolwiek, co nie jest optymalne, a co może odwoływać się do ważnych dla człowieka odczuć, przeżyć i doświadczeń. Socjologia „myślących maszyn” już na starcie eliminuje jej ludzką adekwatność poprzez nietrafną definicję człowieczeństwa. Z pewnością przyszłość sztucznej inteligencji i powstających o niej nauk pójdzie w dwóch, nie do końca współmiernych kierunkach. Jeden skoncentruje się na jej możliwościach technicznych, funkcjonalnych i systemowych. Drugi polegać będzie na poszukiwaniu form nie tylko naśladowania, ale także generowania swoich produktów maksymalnie podobnych do parametrów świata ludzkich zachowań i wartości. Problem jednak w tym, że nie wszystkie one są w stanie poddać się najbardziej nawet złożonej parametryzacji, algorytmizacji i analityce procesów ich integracji. Tak dzieje się z fenomenem ludzkiej, w tym wspólnotowej, tożsamości.

JAKIE TOŻSAMOŚCI?

Świadomość stanowienia odrębnej i spójnej całości kulturowej o własnym systemie wartości rdzennych to niewątpliwie ważne wyróżniki tożsamości społecznej podmiotów wyrastających ponad indywidualną aktywność swoich uczestników¹⁰. Kategoria tożsamości nie jest jednak czymś jednorodnym. Istotne różnice dzielą tożsamość osobową od tożsamości wspólnot, tożsamość roli od tożsamości strukturalnej konkretnych działań i organizacji. Warto więc zwrócić uwagę na istotne wyróżniki takiego typu tożsamości w kontekście hipotetycznej socjologii „myślących maszyn” w jakkolwiek metaforycznym ich ujęciu.

Po pierwsze, tożsamość społeczna jest wyrazem nie tylko formalnej strukturyzacji, ale przede wszystkim podmiotowości wspólnot. Jak wskazywał już jeden z polskich i europejskich prekursorów personalizmu społecznego, były rektor Akademii Krakowskiej, wybitny przedstawiciel humanizmu krakowskiego piętnastego wieku, Paweł Włodkowic, wspólnota przejmując przywołane już atrybuty tożsamości osobowej: godność, wolność, odpowiedzialność¹¹.

¹⁰ Zob. L. D y c z e w s k i, *Naród podmiotem kultury*, w: *Tożsamość polska i otwartość na inne społeczeństwa*, red. L. Dyczewski, Redakcja Wydawnictw KUL, Lublin 1996, s. 11-38.

¹¹ Zob. S. W i e l g u s, *Z badań nad średniowieczem*, Redakcja Wydawnictw KUL, Lublin 1995; L. E h r l i c h, *Paweł Włodkowic i Stanisław ze Skarbimierza*, Państwowe Wydawnictwo Naukowe, Warszawa 1954; E. A. W e s o ł o w s k a, *Paweł Włodkowic – współczesne znaczenie poglądów*

Wartości te nie mogą istnieć w żaden sztuczny, pozorowany, półrealny i wykoncypowany sposób. Tak długo, jak sztuczna inteligencja będzie jedynie naśladować człowieka i tworzone przez niego byty, zawsze skazana będzie na fundamentalne uproszczenia, choćby w kwestii praktykowania wartości autotelicznych. Przykładem mogą być działania heroiczne, których rezultatu nie da się zaplanować czy optymalizować. Nie mieszczą się one w strategiach socjotechniki zarządzania¹². Sztuczna inteligencja nie zaplanowałaby ani powstania warszawskiego, ani obecnej bohaterskiej obrony Ukrainy, której podmiotowość kontrastuje z oczekiwaniami poddania się i rezygnacji z własnej suwerenności. Są to, jak widać, wymogi, którym nie może sprostać nie tylko sztuczna, ale i naturalna inteligencja wielu europejskich polityków i intelektualistów, robiących wszystko, aby upodobnić się do tej pierwszej.

Po drugie, tożsamość wspólnot jest dobrym przykładem nieredukowalnych potencjałów stymulujących rozwój kompetencji poznawczych i aksjologicznych swoich członków. Przejawiają się one w postawach i działaniach zbiorowych, które przeciwdziałają procesom atomizacji, rozpadu i atrofii społecznych¹³. Te ostatnie prowadzą prostą drogą do destrukcji, wewnętrznej polaryzacji, a nawet wojen kulturowych, powołując się na wartości demokratyczne o własnej i nader często partykularnej interpretacji. W sposób naturalny zachowania te są równoważone przez czynniki integrujące, które zawarte są w potencjale rozwojowym każdej wspólnoty, utrzymując jej podmiotowość, spójność i prawa kulturowe do samodzielnie określanego wzrostu¹⁴.

Po trzecie, tożsamość społeczna, integrując zatomizowane i funkcjonalne identyfikacje zarówno jednostek, jak i wspólnoty, wynika z aktywności zbiorowej jaźni, a więc z tego, co Antonina Kłoskowska nazywała czynnością walencji, odróżniającą wartości rdzenne od peryferyjnych wspólnoty¹⁵. Efektem walidacji jest konkretna konfiguracja i hierarchizacja wartości. Tak budowana tożsamość jest jednocześnie źródłem najwyższego z opisywanych wcześniej typów więzi oraz swoistej wrażliwości spełniającej rolę azymutu

i dokonań, Wydawnictwo Naukowe Novum, Płock 2001; *Jagiellonian Ideas Towards Challenges of Modern Times*, red. L. Korporowicz, S. Jaskuła, M. Stefanović, P. Plichta, Jagiellonian Library, Kraków 2017.

¹² Por. T. Kobierzycki, *Osoba – dylematy rozwoju. Studium metakliniczne*, Wydawnictwo Pomorze, Bydgoszcz 1989, s. 201-217.

¹³ Zob. N. Luhmann, *Systemy społeczne. Zarys ogólnej teorii*, tłum. M. Karczmarczyk, Zakład Wydawniczy Nomos, Kraków 2007; F. Braudel, *Gramatyka cywilizacji*, tłum. H. Igalson-Tygielska, Oficyna Naukowa, Warszawa 2006.

¹⁴ Zob. P. Ścigał, *Tożsamość narodowa. Zarysy problematyki*, Księgarnia Akademicka, Kraków 2012.

¹⁵ Por. A. Kłoskowska, *Kultury narodowe u korzeni*, Wydawnictwo Naukowe PWN, Warszawa 2005, s. 39, 103-112.

aksjologicznego¹⁶. Jego obecność określa faktyczny stan zasobów aksjonormatywnych grup i wspólnot oraz umiejętności ich wykorzystania. Są to wymogi tak odległe od znanych nam współcześnie sztucznych systemów inteligencji, że pozwalają wręcz mówić o ich aksjologicznym redukcjonizmie, kreującym iluzoryczne wyobrażenie kompetencji kulturowych zniewolonych mechaniką symbolicznych algorytmów.

Dynamika tworzenia tożsamości daleka jest od kumulacji cząstkowych identyfikacji i układania ich w zaprojektowane wzory¹⁷. Procesy walidacyjne nie określają socjalizacyjnej mechaniki, a nawet ją relatywizują w sytuacjach zaistniałych konfliktów, sprzeczności i aksjologicznych dylematów. Dlatego są one według Kłoskowskiej elementem ludzkiej wrażliwości i gotowości do radykalnej krytyki skierowanej przeciw sobie. Nadzieja reintegracji tożsamości wyrasta bowiem z naczelnej zasady ukierunkowania ku przyszłości, bogactwa potencjałów rozwojowych odkrywanych poza zasadą ekstrapolacji, przewidywalności, optymalizacji i teorii racjonalnych wyborów¹⁸. Rozwajowe prawo człowieka nie mieści się w żadnej logice systemowo rozumianej inteligencji. Jest nim prawo do szaleństwa, buntu i kontestacji, a tego nie wytrzyma najbardziej złożona maszyna, choć w buncie tym może być ukryta niejedna logika, dynamika i nieznana jeszcze prawda. Logika rozwoju tożsamości nie ma charakteru liniowego, nader często dokonuje się poprzez etapy dezintegracji struktur silnie ukształtowanych i znacznego przewartościowania kluczowych komponentów. Analogiczna logika sztucznej inteligencji musiałaby więc zawierać możliwości własnej dezintegracji, zagubienia i ponoszenia wszystkich tego konsekwencji. Kazimierz Dąbrowski, jeden z najwybitniejszych polskich znawców procesów rozwojowych osobowości, twórca teorii dezintegracji pozytywnej, z naciskiem i analityczną starannością udokumentował negatywne aspekty statycznych stanów integracji, szczególnie tak zwanej integracji pierwotnej, oraz twórcze potencjały ich rozbicia. Istotnym elementem teorii Dąbrowskiego było odróżnienie faz czy etapów rozwojowych, które mają swoje naturalne i następujące po sobie sekwencje, od poziomów rozwoju¹⁹. Te ostatnie daleko wybiegają poza kognitywne i racjonalno-analityczne umiejętności dojrzewającej osobowości. Angażują takie przymioty, jak autorefleksja, niezadowolenie z siebie, poczucie niespełnienia, a nawet zawstydzenie

¹⁶ Zob. H.C. White, *Tożsamość i kontrola. Jak wylaniają się formacje społeczne*, tłum. A. Hałas, Zakład Wydawniczy Nomos, Kraków 2011.

¹⁷ Por. T. Kobierzycki, *Jaźń i tożsamość. Studia z filozofii człowieka*, Wydawnictwo Muza, Warszawa 2012, s. 42-55.

¹⁸ Por. A. Kłosowska, *Socjologia kultury*, Państwowe Wydawnictwo Naukowe, Warszawa 1981, s. 172-215.

¹⁹ Por. K. Dąbrowski, *Dezintegracja pozytywna*, Państwowy Instytut Wydawniczy, Warszawa 1979, s. 46-63.

sobą oraz odróżnienie siebie realnego i możliwego. Wykazują one relatywizacje, a czasem wręcz odwrócenie dotychczasowych kryteriów wzrastania, eksponując ich aksjologiczny charakter. Ich walidacja, także w odniesieniu do rzeczywistości społecznej, dystansuje się od czynników wzrostu, postępu i adaptacji²⁰. Odwołanie się do kryterium poziomów rozwoju odróżnia psychologię rozwojową, badającą prawidłowości w sekwencji etapów życia człowieka, od psychologii rozwoju, respektującej kategorie egzystencjalne i walidacje ważne dla osobowych wymiarów rozwoju²¹. Różnice te w znacznej mierze znajdują swoje korelaty w rozwoju tożsamości, w tym tożsamości kulturowej wspólnot²². Socjologia sztucznej inteligencji rozróżnień takich, na obecnym etapie jej rozwoju, po prostu nie potrzebuje, gdyż nie może zoperacjonalizować wielu kluczowych dla niej pojęć, co oznacza, że różnic tych nie widzi w praktyce i w znacznej mierze rozmija się z przedmiotem właściwie pojętej socjologii.

Pytając o możliwości socjologii „myślących maszyn” w analizie najbardziej złożonych procesów, które poprzez systemy relacji i budowanych przez nie więzi prowadzą do wykształcenia zarówno indywidualnych, jak i ponadindywidualnych tożsamości, należy postawić pytanie najtrudniejsze: w jakim stopniu świat rzeczywistości konceptualnej nadchodzącej ery reprezentacji cyfrowych odpowie na wyzwania logotwórczych dynamizmów kultury, a więc tych, które odkrywają, praktykują i przetwarzają potrzebę doświadczania sensu jako podstawowego czynnika trwania oraz rozwoju osób i wspólnot?²³ Problemu nie wyczerpuje przy tym tylko jedna kategoria „sensu”. Możliwe jest wyodrębnienie przynajmniej ośmiu grup znaczeniowych pojęcia „logosu” ukazując dynamikę nie tyle „pola”, co „przestrzeni semantycznej pojęcia”²⁴. Część z tych grup, a przynajmniej fragmenty ich zakresów, mogą stać się przedmiotem działania sztucznej inteligencji, na przykład umiejętność rozpoznawania relacji, zdolność do rozumnego „oglądu świata” oraz szczególny rodzaj kompetencji do rozpoznawania i klasyfikacji problemów, choć przyznanie jej takich umiejętności czy też władz sądenia

²⁰ Por. J.S. P a s i e r b, *Pionowy wymiar kultury*, Wydawnictwo Znak, Kraków 1983, s. 55. „W człowieku natura wyrasta najwyżej i przekracza prawa ziemskiego ciężenia. Uświadamiamy sobie grozę pozbawienia kultury ludzkiej tego co znajduje się powyżej dachu owego zaplecza metafizycznego filozofii i etyki religii” (tamże).

²¹ Por. T. K o b i e r z y c k i, *Pojęcie charakteru narodowego i problem człowieczeństwa (w koncepcji Kazimierza Dąbrowskiego)*, „Humanistyka i Przyrodoznawstwo” 14(2008), s. 32-50.

²² Zob. L. K o r p o r o w i c z, *Przemiany kultury narodowej w świetle teorii dezintegracji pozytywnej Kazimierza Dąbrowskiego*, „Polska Myśl Pedagogiczna” 4(2018), s. 191-208.

²³ Zob. t e n ż e, *Logotwórcze dynamizmy kultury*, „Studia Socjologiczne” 1989, nr 3, s. 109-131; W. S t r ó ż e w s k i, *Istnienie i sens*, Wydawnictwo Znak, Kraków 1994.

²⁴ Por. L. K o r p o r o w i c z, *Osobowość i komunikacja w społeczeństwie transformacji*, Instytut Kultury, Warszawa 1996, s. 107-112.

byłoby dalece umowne i zdecydowanie przesadne²⁵. Są jednak takie grupy znaczeń wskazujących na przymioty, których od „myślących maszyn” nikt wręcz nie oczekuje, jak zdolność rozpoznawania wartości (w tym autotelicznych), wewnętrzne rozważanie duszy, wola wypełniania sensu działań i istnienia, gotowość i sztuka podejmowania dialogu, czy też kontemplowanie Absolutu w łączności z boskim porządkiem świata²⁶.

Redukcje, których musielibyśmy dokonać, czyniłyby z każdego typu tożsamości człowieka jego mocno zniekształcony wizerunek, niemal cień tego, co faktycznie atrybuty tej tożsamości reprezentują, czym są i jak w nim istnieją. Problem logotwórczych dynamizmów kultury nie kończy się bynajmniej na przyznaniu ich obecności, konsekwencji ich działania i adekwatności nauk, które je ignorują. Ich humanistyczne wersje, na przykład przywołana psychologia rozwoju, ale także socjologia humanistyczna i analogicznie zorientowane studia kulturowe, odkrywają znaczenie sygnalizowanej perspektywy rozwojowej i zdecydowanie wymagają jej uwzględnienia, jeśli nie mają one jedynie symulować zdolności wejrzenia w złożoność prawdy o człowieczeństwie i kulturze, a sztuczną inteligencję traktować jako wyraz jej doskonalenia. Każdy z wyodrębnionych obszarów logotwórczych dynamizmów kultury posiada bowiem odmienny poziom rozwoju, który tworzy równie odrębną rzeczywistość. Charakterystyka poszczególnych poziomów zaproponowana przez przywołaną teorię Dąbrowskiego ujawnia radykalne wręcz różnice w obrębie tych samych dynamizmów. Są one niestety pomijane w większości teorii socjologicznych i z pewnością nie dają się rozpoznać za pomocą kluczowych dla sztucznej inteligencji kryteriów wartościowania, odwołujących się do sygnalizowanych już zasad optymalizacji, funkcjonalnej adaptacji, postępu i ekstrapolacji trendów.

Przykładem zasadniczych różnic może być zdolność i warunkowany przez nią dynamizm rozpoznawania wartości. Pierwszy z wyodrębnionych w teorii dezintegracji pozytywnej poziomów – określony jako poziom integracji pierwotnej – jest w istocie poziomem automatyzmów funkcjonalnych procesu socjalizacji, polegającym na wykształceniu powtarzalnych schematów rozpoznawania i realizacji stabilnego kanonu wartości. O ile jego ukształtowanie, a nawet instytucjonalizacja, jest przejawem dojrzewania i wypracowania oznak odrębności oraz podmiotowości kulturowej zarówno osób, jak i wspólnot, o tyle zakodowanie ich kanonu w postać zamkniętych konwencji grozić może

²⁵ Zob. K. P o p i e l s k i, *Logoteoria i logoterapia w kontekście psychologii współczesnej. Człowiek – pytanie otwarte. Studia z logoteorii i logoterapii*, red. K. Popielski, Redakcja Wydawnictw KUL, Lublin 1987.

²⁶ Por. J. L e g o w i c z, *Historia filozofii starożytnej Grecji i Rzymu*, Państwowe Wydawnictwo Naukowe, Warszawa 1973, s. 55; *Słownik pisarzy antycznych*, red. A. Świderkówna, Wiedza Powszechna, Warszawa 1982.

jego alienacją. Konformizm, bierność i reprodukcja postaw to konsekwencje kosztowności wrażliwości etycznej, która zamienia się w rodzaj aksjologicznego algorytmu oraz praktykę przesocjalizowanych doktryn wychowawczych. Wcześniej czy później stają się one elementem represywnych stosunków władzy. Być może pozwalają one na zachowanie porządku społecznego, ale nie są gotowe na jego ewolucję i reagowanie na potrzeby transformacji powodowane tak czynnikami wewnętrznymi, jak i zewnętrznymi. Umiejętność rozpoznawania wartości popada w poważny kryzys. Czy systemy „myślących maszyn” mogą być na tyle wrażliwe, empatyczne, ale i analityczne, aby móc potrzeby te zdiagnozować, a tym samym zaprojektować konieczne zmiany i wytyczyć drogi ich realizacji? Prawdziwy proces rozwojowy musi otworzyć się na uwolnienie od automatyzmów funkcjonalnych, niezdolnych do podejmowania wyzwań interakcyjnych, szczególnie w czasach mobilności, pluralizacji i spotkania kultur. Niezbędne staje się więc poddanie ich dezintegracji. Taką charakterystyką cechuje się drugi poziom w ewentualnej sekwencji działań transgresyjnych. Niestety posiada on wiele dramatycznych i dysfunkcyjnych aspektów, przed którymi sztuczna inteligencja ma wszak ochraniać i skutecznie je eliminować. Dekonstrukcja spetryfikowanych tożsamości to zadanie dalekie od kalkulacji, socjotechniki zarządzania zmianą i algorytmu racjonalnych wyborów. Procesy dezintegracyjne są jednak konieczne w perspektywie logiki rozwoju, która odbiega od logiki wzrostu i pomnażania, jak też od logiki dezintegracyjnej prewencji.

W celu pokonania ograniczenia zarówno poziomu pierwszego, jak i drugiego konieczne jest odkrycie i animacja dynamizmów rozwojowych poprzez ich aktywizację na trzech kolejnych poziomach. Umiejętność rozpoznawania wartości wyrasta na poczuciu istotnego egzystencjalnie braku, lęku i dysonansu między tym, co automatyzmy funkcjonalne zniszczyły, a tym, co przetrwało w głębokiej warstwie aksjologicznej wrażliwości. Trzeci poziom rozwoju osobowości Dąbrowski nazywa dezintegracją wielopoziomową i spontaniczną, co w praktyce oznacza poziom ujawnienia ważnej potrzeby sensu i przywrócenia realizmu istotnych życiowo wartości. Udana witalizacja poziomu trzeciego prowadzi do kolejnego, który cechuje się zaawansowaniem czynników refleksyjnych, samowychowaniem, autoterapią, pokonywaniem determinizmów biologicznych i społecznych oraz umiejętnością autonomicznego projektowania oraz komponowania wartości ponadfunkcjonalnych. To poziom rozwoju autonomicznego, w którym logotwórcze dynamizmy kultury przenikają się i ponownie integrują, wskazując na inspirujące ich ideały poziomu najwyższego. Harmonizuje on i syntetyzuje wszystkie dynamizmy rozwojowe, to, „co jest”, spaja z tym, „co być powinno” i aktywizuje dynamizm autentyzmu, zawierający cztery podstawowe komponenty, takie jak: (1) wszechstronność rozwoju wszystkich aspektów życia wewnętrznego, (2) zaufanie do samodziel-

nie określonej drogi realizacji własnego ideału osobowości, (3) świadomość odrębności siebie przy uznaniu odrębności innych, (4) rozwiązanie relacji między „ja” a „ty”, odwołujące się do wysoko rozwiniętych jakości tego, co indywidualne, i tego, co wspólne, do odpowiedzialności zarówno w stosunku do siebie, jak i innych ludzi²⁷.

Jak widać tożsamość – tak w wersji osobowej, jak i wspólnotowej – wykazuje się nie tylko odmiennością swoich obszarów, ale także źródeł i potencjałów jej tworzenia. Umiejętność dostrzeżenia ich różnorodności z uwzględnieniem kryteriów dalekich od instrumentalności, a czasami w stosunku do nich wręcz dysfunkcyjnych stanowi o wiarygodności i adekwatności analizujących ją nauk²⁸. Dotyczy to również ich obietnic, które rysują się na horyzoncie nie tylko sztucznej inteligencji, ale także sztucznej kultury, hybrydalnego i równie sztucznego człowieka. Zachwyt serią potrzebnych udogodnień i poszerzenie sfery działań człowieka nie może zasłonić rozumienia tego, kim jest, szczególnie wtedy, gdy nie okazuje się doskonały.

*

Wielopoziomowość logotwórczych dynamizmów kultury, szczególnie w odniesieniu do procesu tworzenia tożsamości osobowej i kulturowej człowieka, wskazuje na konsekwencje redukcji jej istotnych cech. Nie tylko w socjologicznym, ale także w humanistycznym rozumieniu tożsamości widać słabość jednowymiarowego, progresywnego ujęcia problemu. Z łatwością można sporządzić listę procesów, umiejętności i wartości, które trudno zidentyfikować w logice funkcjonowania „myślących maszyn” i najbardziej nawet metaforycznie pojmowanej inteligencji bez jej wymiaru emocjonalnego, kulturowego, a nawet moralnego. Być może dlatego bardziej wiarygodne będzie nakreślenie projektu socjologii działań wspomaganych systemem „myślących maszyn”, bo te zabierać będą w życiu społecznym coraz większą część naszej uwagi, nasilając ich wykorzystanie. Z pewnością modyfikacji ulegnie wiele ludzkich zachowań, a nawet postaw, ale nie oznacza to, iż staną się one wyznacznikiem rozumienia i rozwoju samego człowieka w każdym wariantcie jego człowieczeństwa. Nauki społeczne i humanistyczne, unikając zbyt daleko posuniętych redukcji, nie mogą problemu tego zastępować zagadnieniami innych nauk, szczególnie technicznych, które anektują coraz bardziej nie tylko

²⁷ Zob. *Theory Building in Developmental Psychology*, red. P. van Geert, North-Holland, Amsterdam 1986.

²⁸ Zob. J. K o c i u b a, *Idea Ja i koncepcji tożsamości. Zmiana znaczenia idei Ja i koncepcji tożsamości w nauce i kulturze*, Wydawnictwo Episteme, Lublin 2014; A. Elliot, *Koncepcje „ja”*, tłum. S. Królak, Wydawnictwo Sic!, Warszawa 2007.

zakres, lecz i cele współczesnych nauk. Praktyka zamienników i mylących metafor wcześniej czy później zagubi ważne poznawcze, ale nade wszystko istotę naszego człowieczeństwa.

BIBLIOGRAFIA / BIBLIOGRAPHY

- A Greek-English Lexicon*. Edited by Henry George Liddell and Robert Scott. Oxford: Clarendon Press, 1968.
- Braudel, Fernand. *Gramatyka cywilizacji*. Translated by Hanna Igalsen-Tygielska. Warszawa: Oficyna Naukowa, 2006.
- Cassirer, Ernst. *Filozofia form symbolicznych*. Translated by Andrzej Karalus and Przemysław Parszutowicz. Kęty: Wydawnictwo Marek Derewiecki, 2018.
- Culture and Society: Contemporary Debates*. Edited by Jeffrey C. Alexander and Steven Seidman. Cambridge: Cambridge University Press, 1990.
- Dąbrowski, Kazimierz. *Dezintegracja pozytywna*. Warszawa: Państwowy Instytut Wydawniczy, 1979.
- . *Zdrowie psychiczne a wartości ludzkie*. Warszawa: Polskie Towarzystwo Higieny Psychiczej, 1974.
- Dudek, Zenon Waldemar, and Andrzej Pankalla. *Psychologia kultury: Doświadczenia graniczne i transkulturowe*. Warszawa: Wydawnictwo Psychologii Kultury, Eneteia, 2008.
- Dyczewski, Leon. “Naród podmiotem kultury.” In *Tożsamość polska i otwartość na inne społeczeństwa*. Edited by Leon Dyczewski. Lublin: Redakcja Wydawnictw KUL, 1996.
- Ehrlich, Ludwik. *Paweł Włodkowic i Stanisław ze Skarbimierza*. Warszawa: Państwowe Wydawnictwo Naukowe, 1954.
- van Geert, Paul. “The Concept of Deyelopment.” In *Theory Building in Developmental Psychology*. Edited by Paul van Geert. Amsterdam [etc.]: North-Holland, 1986.
- Elliot, Anthony. *Koncepcje „ja”*. Translated by Sławomir Królak. Warszawa: Wydawnictwo Sic!, 2007.
- Golka, Marian. *Cywilizacja współczesna i globalne problemy*. Warszawa: Oficyna Naukowa, 2012.
- Granat, Wincenty. *Personalizm chrześcijański: Teologia osoby ludzkiej*. Sandomierz: Wydawnictwo Diecezjalne i Drukarnia w Sandomierzu, 2018.
- Habermas, Jurgen. *Teoria działania komunikacyjnego*. Vol. 1. *Racjonalność działania a racjonalność społeczna*. Translated by Andrzej Maciej Kaniowski. Warszawa: Państwowe Wydawnictwo Naukowe, 1999.
- Handbook of Sociology*. Edited by Neil J. Smelser. London and New Delhi: Sage Publication Newbury Park, 1988.
- Jagiellonian Ideas Towards Challenges of Modern Times*. Edited by Leszek Korporowicz, Sylwia Jaskuła, Malgożata Stefanovič, and Paweł Plichta. Kraków: Jagiellonian Library, 2017.

- Kłoskowska, Antonina. *Kultury narodowe u korzeni*. Warszawa: Wydawnictwo Naukowe PWN, 2005.
- . *Socjologia kultury*. Warszawa: Państwowe Wydawnictwo Naukowe, 1981.
- Kobierzycki, Tadeusz. *Jaźń i tożsamość: Studia z filozofii człowieka*. Warszawa: Wydawnictwo Muzaios, 2012.
- . *Osoba – dylematy rozwoju: Studium metakliniczne*. Bydgoszcz: Wydawnictwo Pomorze, 1989.
- . “Pojęcie charakteru narodowego i problem człowieczeństwa (w koncepcji Kazimierza Dąbrowskiego).” *Humanistyka i Przyrodoznawstwo*, no. 14 (2008): 29–50.
- Kociuba, Jolanta. *Idea Ja i koncepcji tożsamości: Zmiana znaczenia idei Ja i koncepcji tożsamości w nauce i kulturze*. Lublin: Wydawnictwo Episteme, 2014.
- Korporowicz, Leszek. “Komunikacja symboliczna w rozwoju tożsamości kulturowej.” In *Tożsamość polska w odmiennych kontekstach*. Edited by Leon Dyczewski and Dariusz Wadowski. Lublin: Wydawnictwo KUL, 2009.
- . “Logotwórcze dynamizmy kultury.” *Studia Socjologiczne*, no. 3 (1989): 109–31.
- . “Przemiany kultury narodowej w świetle teorii dezintegracji pozytywnej Kazimierza Dąbrowskiego.” *Polska Myśl Pedagogiczna*, no. 4 (2018): 191–208.
- . “Źródła i typy redukcjonizmów we współczesnych analizach kultury.” *Roczniki Kulturoznawcze*, no. 7 (2016): 5–30.
- Kozielecki, Józef. *Koncepcja transgresyjna człowieka*. Warszawa: Państwowe Wydawnictwo Naukowe, 1987.
- Legowicz, Jan. *Historia filozofii starożytnej Grecji i Rzymu*. Warszawa: Państwowe Wydawnictwo Naukowe, 1973.
- Luhmann, Niklas. *Systemy społeczne: Zarys ogólnej teorii*. Translated by Michał Karczmarczyk. Kraków: Zakład Wydawniczy Nomos, 2007.
- Matsumoto, David, and Linda Juang. *Psychologia międzykulturowa*. Gdańsk: Gdańskie Wydawnictwo Psychologiczne, 2007.
- Palczyński, Tadeusz. *Relacje międzykulturowe w dobie kryzysu ideologii i polityki wielokulturowości*. Kraków: Księgarnia Akademicka, 2017.
- Pasierb, Janusz Stanisław. *Pionowy wymiar kultury*. Kraków: Wydawnictwo Znak, 1983.
- Pisarek, Walery. *Wstęp do nauki o komunikowaniu*. Warszawa: Wydawnictwa Akademickie i Profesjonalne, 2008.
- Popielski, Krzysztof. “Logoteoria i logoterapia w kontekście psychologii społecznej.” In *Człowiek – pytanie otwarte: Studia z logoteorii i logoterapii*. Edited by Kazimierz Popielski. Lublin: Redakcja Wydawnictw KUL, 1987.
- Słownik pisarzy antycznych*. Edited by Anna Świderkówna. Warszawa: Wiedza Powszechna, 1982.
- Smolicz, Jerzy. *Współkultury Australii*. Warszawa: Oficyna Naukowa, 1999.
- Stróżewski, Władysław. *Istnienie i sens*. Kraków: Wydawnictwo Znak, 1994.
- Ścigaj, Paweł. *Tożsamość narodowa: Zarysy problematyki*. Kraków: Księgarnia Akademicka, 2012.

- Trąbka, Agnieszka. *Tożsamość rekonstruowana: Znaczenie migracji w biografach Third Culture Kids*. Warszawa: Wydawnictwo Naukowe Scholar, Warszawa 2014.
- Wesołowska, Eugenia Anna. *Paweł Włodkowic – współczesne znaczenie poglądów i dokonań*. Płock: Wydawnictwo Naukowe Novum, 2001.
- White, Harrison C. *Tożsamość i kontrola: Jak wylaniają się formacje społeczne*. Translated by Agnieszka Hałas. Kraków: Zakład Wydawniczy Nomos, 2011.
- Wielgus, Stanisław. *Z badań nad średniowieczem*. Lublin: Redakcja Wydawnictw KUL, 1995.

ABSTRAKT / ABSTRACT

Leszek KORPOROWICZ – Redukcyjne pułapki socjologii „myślących maszyn”
DOI 10.12887/36-2023-3-143-13

Najistotniejszym celem artykułu jest wskazanie na rozliczne uproszczenia, które dokonywane są w analizie rzeczywistości społecznej, wspomaganej współczesnymi i przyszłymi zastosowaniami sztucznej inteligencji. Wykonanie tego zadania wymaga omówienia zagadnienia redukcji, na które skazana jest hipotetyczna socjologia „myślących maszyn” jako typowy przykład nieporozumień wywołanych częstym tworzeniem zbyt daleko idących metafor. Co ważniejsze, konsekwencją tego typu praktyk jest ukryta zamiana właściwego przedmiotu nauk społecznych i humanistycznych, którym jest człowiek i jego człowieczeństwo, na koncentrację wokół używanych przez nie instrumentów, czego przykładem jest tak zwana humanistyka cyfrowa. Analizy odniesione będą do trzech obszarów badawczych socjologii, takich jak: relacje, więzi i tożsamości społeczne w kontekście możliwości, które spełniać miałyby socjologia zaawansowanych technologicznie maszyn, co nie jest równoznaczne z badaniem ich wpływu na działalność człowieka. W najbardziej istotnej kwestii procesu budowania oraz typologii tożsamości społecznych wykorzystana zostanie koncepcja logotwórczych dynamizmów kultury, jako ważnego czynnika tworzenia sensu i wartości ludzkich, który trudno odnaleźć w funkcjonalnej, instrumentalnej i parametryzowanej rzeczywistości maszyn oraz ich systemów. Co więcej, w artykule zostanie wykazana radykalna różnica każdego z dynamizmów w zależności od poziomu ich realizacji, na którą wskazuje teoria dezintegracji pozytywnej Kazimierza Dąbrowskiego i socjologii kulturalistycznej Antoniny Kłoskowskiej. W rezultacie powstanie cała lista wymogów, na które socjologia sztucznej inteligencji odpowiedzieć nie może, o ile porzuci fundamentalne kategorie rzeczywistości ludzkiej, jak godność, honor, wolność czy odpowiedzialność, a więc wartości autoteliczne człowieka.

Słowa kluczowe: redukcjonizm aksjologiczny, sztuczna inteligencja, logotwórcze dynamizmy kultury, socjologia humanistyczna, koncepcja rozwoju wielopoziomowego

Kontakt: Zakład Humanistycznych Problemów Nauk o Kulturze, Instytut Studiów Międzykulturowych, Wydział Studiów Międzynarodowych i Politycznych, Uniwersytet Jagielloński, ul. Reymonta 4, 30-059 Kraków
E-mail: leszek.korporowicz@uj.edu.pl
<https://ism.uj.edu.pl/leszek-korporowicz>
ORCID: 0000-0002-5890-5991

Leszek KORPOROWICZ, Reductive Pitfalls of the Sociology of “Thinking Machines”
DOI 10.12887/36-2023-3-143-13

The most important objectives of the article is to indicate the numerous simplifications that are made in the analysis of social reality supported by modern and future applications of artificial intelligence. The performance of this task is based on a discussion of the reductions to which the hypothetical sociology of “thinking machines” is doomed as a typical example of misunderstanding caused by the frequent creation of too far-reaching metaphors. More importantly, the consequence of this type of practice is the hidden replacement of the proper subject of social sciences and humanities, which is man and his humanity, with the instruments they use, an example of which is the so-called “digital humanities.” The analyses will relate to three research areas of sociology, such as relationships, bonds and social identities in the context of the possibilities that the sociology of technologically advanced machines and their impact on humans would meet. In the most important issue of the process of building and the typology of social identities, the article refers to the concept of logo-creating dynamics of culture as an important factor in forming the meaning and value of human actions, which is difficult to find in the functional, instrumental and parameterized reality of machines and their systems. Moreover, the article will show a radical difference of each of the dynamics depending on the level of their implementation, which is indicated by the theory of positive disintegration by Kazimierz Dąbrowski and cultural sociology by Antonina Kłoskowska. As a result, a list of requirements is created to which the sociology of artificial intelligence cannot respond if it abandons the fundamental categories of human reality, such as dignity, honor, freedom or responsibility, and thus the autotelic values of man.

Keywords: axiological reductionism, artificial intelligence, logo-creative dynamics of culture, humanistic sociology, the concept of multi-level development

Contact: Zakład Humanistycznych Problemów Nauk o Kulturze, Instytut Studiów Międzykulturowych, Wydział Studiów Międzynarodowych i Politycznych, Uniwersytet Jagielloński, ul. Władysława Reymonta 4, 30-059 Cracow, Poland
E-mail: leszek.korporowicz@uj.edu.pl
<https://ism.uj.edu.pl/leszek-korporowicz>
ORCID: 0000-0002-5890-5991

Renata TOMASZEWSKA

RAJ – PIEKŁO – TRIUMF Sztuczna inteligencja w scenariuszach przyszłości

W optyce „rajskiej” ludzkość ma zostać diametralnie udoskonalona dzięki postępowi w zakresie sztucznej inteligencji. W wizji „piekielnej” przewiduje się katastroficzną zagładę, oznaczającą kres człowieka w znanej dziś postaci. Trzeci scenariusz „triumfalny” jest bardziej skomplikowany, niejednoznacznie zdeterminowany. Według tej wizji to nadal ludzie, a nie AI czy AGI, będą wybierać kształt własnej przyszłości, kierując się wyznawanymi wartościami.

Oddziaływanie technologii na rozwój ludzkości, w tym postęp w dziedzinie sztucznej inteligencji, wymaga ze wszech miar analiz humanistycznych, w tym filozoficznych. Problematyka ta inspiruje reprezentantów także innych dziedzin i dyscyplin naukowych. Rozwój sztucznej inteligencji nie pozostaje bez wpływu na formułowanie prognoz dotyczących przyszłości człowieka. Człowieczeństwo to przedmiot między innymi refleksji pedagogicznej. „Jak określić i zdefiniować człowieczeństwo? To zadanie filozofii. Jak je realizować i jak je budować w człowieku? To zadanie pedagogiki”¹.

W niniejszym artykule zaprezentowano perspektywę pedagoga; wyjaśniono rozumienie pojęć „sztuczna inteligencja” i „superinteligencja” oraz ukazano wybrane hipotetyczne scenariusze przyszłości. Wśród nich wskazano na triadę: Raj – Piekło – Triumf², które można potraktować jako metafory interpretacji rozwoju technologicznego, przydatne zarówno dla filozofii, jak i pedagogiki.

SZTUCZNA INTELIGENCJA I SUPERINTELIGENCJA

Począwszy od wczesnych rozważań filozoficznych, przez odkrycia psychologii, po współczesne metody neuronauki, człowiek próbuje zrozumieć sposób, w jaki postrzega otaczającą go rzeczywistość i w niej funkcjonuje. Wiedza o własnym poznaniu stwarza pokusę próby zbudowania struktur, które

¹ W. Furmank, *Człowiek jako obiekt badań humanistycznej pedagogiki pracy*, „Labor et Educatio” 2014, nr 2, s. 16.

² Zob. J. Garreau, *Radykalna ewolucja. Czy człowiek udoskonalony przez naukę i technikę będzie jeszcze człowiekiem?*, tłum. A. Kłoch, A. Michalski, Wydawnictwo Prószyński i S-ka, Poznań 2005.

działałyby na wzór ludzi, a może nawet lepiej. Idea ta stanowi podstawę badań nad sztuczną inteligencją (SI; ang. artificial intelligence, AI)³.

W swoich dążeniach do stworzenia struktur inteligentnych człowiek stara się projektować rozwiązania, które możliwie najwierniej odwzorowywałyby to, w jaki sposób myśli i działa. Modelowanie systemów sztucznej inteligencji⁴ tak, aby działały inspirowane funkcjonowaniem ludzkiego mózgu, jest krokiem do stworzenia rozwiązań „zachowujących się tak, jak człowiek”. Projektanci systemów sztucznej inteligencji opracowują metody symulacji poszczególnych zdolności umysłowo-poznawczych człowieka i odnoszące się do nich konstrukty. Obszary zastosowań sztucznej inteligencji obejmują: percepcję i rozpoznawanie obrazów, reprezentację wiedzy (to znaczy pozyskiwanie wiedzy o otoczeniu, jej zapamiętywanie w postaci umożliwiającej szybką, adekwatną reakcję na bodźce płynące z tego środowiska), rozwiązywanie problemów, wnioskowanie, podejmowanie decyzji, planowanie, przetwarzanie języka naturalnego, uczenie, manipulację i lokomocję, a nawet inteligencję społeczną, emocjonalną oraz kreatywność (na przykład wyrażanie emocji poprzez zmiany mimiki mechanicznej twarzy, rozpoznawanie nastroju i intencji człowieka na podstawie jego mowy, generowanie utworów muzycznych i plastycznych)⁵.

Wymienione obszary „kompetencji” systemów inteligentnych stanowią podwaliny współczesnych metod i technologii sztucznej inteligencji, której przypisuje się dwa podstawowe znaczenia:

(1) hipotetycznej inteligencji realizowanej w procesie inżynierskim, czyli w warunkach sztucznych;

(2) technologii i dziedziny badań naukowych informatyki znajdującej się na styku z neurobiologią, psychologią, filozofią i kognitywistyką, której za-

³ Nim ludzkość zyskała AI mieliśmy „cybernetykę” – ideę automatycznego i samoregulującego się sterowania wyłożoną w pracach Norberta Wienera *Cybernetics or Control and Communication in the Animal and the Machine* (zob. N. W i e n e r, *Cybernetics or Control and Communication in the Animal and the Machine*, Wiley, New York 1948) oraz *The Human Use of Human Beings* (zob. t e n ż e, *The Human Use of Human Beings*, Houghton Mifflin, Boston 1950). Wiener wyraził w nich swoje obawy przed niekontrolowanymi i nieprzewidywalnymi konsekwencjami pojawienia się nowych technologii. Natomiast to pionier komputerów – amerykański informatyk i laureat Nagrody Turinga (1971) – John McCarthy zaproponował termin „sztuczna inteligencja” i stał się założycielem tej dziedziny (por. J. B r o c k m a n, *Wprowadzenie. O nadziejach i pułapkach związanych z AI*, w: *Człowiek na rozdrożu. Sztuczna inteligencja – 25 punktów widzenia*, red. J. Brockman, tłum. M. Machnik, Wydawnictwo Helion, Gliwice 2020, s. 9, 13).

⁴ Jednym z podstawowych źródeł sporu dotyczącego definicji sztucznej inteligencji wydają się różnice w sposobie określania samego pojęcia „inteligencja” oraz terminów z nim związanych, takich jak: umysł, poznanie, wiedza. Są one przedmiotem zainteresowania filozofii, w tym filozofii poznania – epistemologii, oraz psychologii (por. M. F l a s i Ń s k i, *Wstęp do sztucznej inteligencji*, Wydawnictwo PWN, Warszawa 2018, s. 217-227).

⁵ Por. tamże, s. 228-240; A. W o d e c k i, *Sztuczna inteligencja w kreowaniu wartości organizacji*, Wydawnictwo Edu-Libri, Kraków–Legionowo 2018, s. 68-70.

daniem jest konstruowanie maszyn i programów komputerowych zdolnych do realizacji wybranych funkcji umysłu i zmysłów, a które nie będą podlegać prostej numerycznej algorytmizacji⁶.

W drugim z przytoczonych znaczeń sztuczna inteligencja rozumiana jest jako nauka o tym, jak produkować maszyny wyposażone w niektóre cechy ludzkiego umysłu. Ze względu na kryterium uniwersalności wyróżnia się:

(a) słabą lub wąską sztuczną inteligencję (ang. weak lub narrow artificial intelligence), która jest stosowana tylko do określonych czynności lub konkretnych typów problemów. Koncentruje się ona na jednym wąskim zadaniu, które potrafi wykonać lepiej od człowieka. Najczęściej występuje w postaci asystentów głosowych (na przykład Cortana czy Siri), automatycznych tłumaczy (Google Translator) czy autonomicznych samochodów (Tesla);

(b) silną lub ogólną sztuczną inteligencję (ang. strong lub general artificial intelligence, AGI), która polegać ma na inteligentnych, dysponujących wszechstronną wiedzą i zdolnościami poznawczymi, systemach, które potrafią samodzielnie myśleć i wykonywać zadania tak samo sprawnie, jak wykonałby je człowiek; także te zadania, których systemy te wcześniej nie znały. Gdyby silna inteligencja istniała – bo jeszcze nie istnieje – byłaby maszyną zdolną do zrozumienia świata i każdego człowieka, posiadającą taką jak ludzie, a z czasem jeszcze doskonalszą zdolność uczenia się i działania. Stałaby się też wówczas superinteligencją⁷.

Sztuczna inteligencja to zatem program, który nie wymaga programowania „wprost”, a który w oparciu o dostarczone dane próbuje odtworzyć (i przewyższyć) działanie ludzkich operatorów w wielu skomplikowanych zadaniach. Jego trzy główne cele to: przyspieszenie powtarzalnych procedur przez zastąpienie ludzi robotami lub maszynami, bycie „wydajniejszym” od ludzkich mózgów przez „uczenie się” i pamięć, rozpoznawanie wzorów oraz podejmowanie decyzji w sposób natychmiastowy i efektywny⁸. W ramach tej nauki czyni się próby zbudowania maszyn, które myślą i wyciągają wnioski na podstawie danych, zamiast operować we względnie ograniczonej przestrzeni o ustalonych z góry procedurach i wynikach⁹. Inteligentne systemy AI rozpoznają wzorce i pamiętają przeszłe wydarzenia, ucząc się z nich. Dzięki temu

⁶ Por. A.K. Przegląd, *Istoty wirtualne. Jak fenomenologia zmieniała sztuczną inteligencję*, Wydawnictwo Universitas, Kraków 2016, s. 239n.

⁷ Zob. hasło „Sztuczna inteligencja”, w: *Słownik*, <https://www.sztuczna-inteligencja.org.pl/definicja/sztuczna-inteligencja>.

⁸ Por. A. Pawlicka, M. Pawlicki, M. Choraś, *Zawód roku 2020. Jak go zdobyć?*, w: *Praca i kariera w warunkach nowego millennium. Implikacje dla edukacji*, red. R. Tomaszewska, Wydawnictwo UKW, Bydgoszcz 2022, s. 145-147.

⁹ Zob. A. Adcock, *These Are the Exact Skills You Need to Get a Job in Artificial Intelligence*, Ladders, 5 II 2020, <https://www.theladders.com/career-advice/these-are-the-exact-skills-you-need-to-get-a-job-in-artificial-intelligence>.

każda kolejna decyzja jest „mądrzejsza”, bardziej logiczna i naturalna. Sztuczna inteligencja daje komputerom możliwość wykonywania takich czynności, jak: rozpoznawanie, diagnoza, planowanie, przewidywanie i tak dalej, bez bycia zaprogramowanymi do tych konkretnych zadań. Ściśle z nią związane jest tak zwane uczenie maszynowe (ang. machine learning, ML). Skupia ono się na tworzeniu algorytmów i modeli statystycznych, które potrafią nauczyć się „rosnąć” i „zmieniać”, gdy natrafią na nowe, nieznane im dotąd dane, stopniowo poprawiając swoje wyniki w kwestii konkretnego zadania. Istnieją cztery główne rodzaje systemów uczenia maszynowego: nadzorowane, nienadzorowane, częściowo nadzorowane i wzmacniane. Różnią się one od siebie radykalnie, między innymi sposobem podania danych i (lub) wzorców, stąd każdy z nich znajdzie odmienny rodzaj zastosowania¹⁰. Uczenie maszynowe jest obecnie jedną z najbardziej pożądanych technologii na świecie¹¹. W istocie trudno wymyślić dziedzinę, w której ML nie znalazłoby zastosowania, co oznacza wzrastające zapotrzebowanie na ekspertów od budowania, testowania i stosowania sztucznej inteligencji w praktyce¹². Współcześnie bowiem sztuczna inteligencja „uczy się jeździć”, „uczy się tłumaczyć”, „uczy się słuchać”, „uczy się stawiać diagnozy”, „uczy się, jak zarabiać pieniądze”, „uczy się

¹⁰ Poszczególne rodzaje uczenia maszynowego zostały zastosowane w kilku transdyscyplinarnych projektach, w których wzięła udział autorka niniejszego artykułu. Dotyczyły one: (1) wykorzystania sztucznych sieci neuronowych do stworzenia narzędzia diagnostycznego stanu równowagi praca–życie pozazawodowe; (2) innowacyjnego badania eksploracji reguł asocjacyjnych dotyczących związków między zmianami w umiejętnościach cyfrowych a świadomością cyberbezpieczeństwa, zachodzącymi w pracy zdalnej podczas pandemii COVID-19; (3) multidyscyplinarnego eksperymentu z wykorzystaniem rozwiązań eksploracji danych dotyczącego rodzajów i wymiarów konfliktu występującego między karierą a macierzyństwem kobiet-naukowczyń (zob. A. Pawlicka, M. Pawlicki, R. Tomaszewska, M. Choraś, R. Gerlach, *Innovative Machine Learning Approach and Evaluation Campaign for Predicting the Subjective Feeling of Work-Life Balance Among Employees*, PLoS ONE 2022, nr 15(5), <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0232771>; A. Pawlicka, R. Tomaszewska, E. Krause, D. Jaroszevska-Choraś, M. Pawlicki, M. Choraś, *Has the Pandemic Made Us More Digitally Literate?*, „Journal of Ambient Intelligence and Humanized Computing” 2022, <https://link.springer.com/article/10.1007/s12652-022-04371-1>; E. Krause, R. Tomaszewska, A. Pawlicka, *Conflicting 'Mother-Scientist' Roles. An Innovative Application of Basket Analysis in Social Research*, PLoS ONE 2022, nr 17(10), <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0276201>).

¹¹ Systemy AI stają się źródłem przewagi technologicznej i konkurencyjnej, stąd przeznaczają się olbrzymie środki finansowe na ich konstruowanie (zob. Communication from the Commission to the European Parliament, the European Council, the Council, *The European Economic and Social Committee and the Committee of the Regions on Artificial Intelligence for Europe*, Brussels, 25.04.2018, COM(2018) 237 final, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN>; European Commission, *White Paper: On Artificial Intelligence: A European Approach to Excellence and Trust*, Brussels, 19.02.2020, COM(2020) 65 final, https://commission.europa.eu/system/files/2020-02/commission-white-paper-artificial-intelligence-feb2020_en.pdf).

¹² Por. Pawlicka, Pawlicki, Choraś, dz. cyt., s. 145-147.

prawa”, „uczy się grać w pokera”, a przede wszystkim „uczy się, jak stać się bardziej inteligentna”¹³.

W nawiązaniu natomiast do zagadnienia superinteligencji to stworzenie systemu, który potrafiłby pracować tak, jak ludzie, i wykazywałby przy tym świadomość swego istnienia, stało się przedmiotem badań wielu zespołów naukowych oraz tych związanych z przemysłem. Można przypuszczać, że jeśli naukowcy zajmujący się AI zdołają opracować działającą AGI, to powstanie maszyna, która dorówna człowiekowi inteligencją. Będzie ona korzystać z pełnego katalogu funkcji, którymi operują dostępne dziś komputery, takimi jak zdolność liczenia i pozyskiwania informacji z prędkością, której my jako ludzie nigdy nie osiągniemy. Większość badaczy przyjmuje dość pesymistyczny scenariusz, że taki supersystem może kiedyś skierować swoją inteligencję nie na świat zewnętrzny, ale na siebie. Skupiając się na poprawie budowy, przeprojektowaniu i wykorzystaniu technik programowania ewolucyjnego, doprowadziłby do samoulepszenia. Spowodowałoby to powtarzalne rekurencyjne ulepszenia. Z każdą poprawką system uczyłby się coraz to nowych rzeczy, aż wreszcie doszłoby do „eksplozji inteligencji”, która sprawiłaby, że musielibyśmy żyć wśród maszyn, które są od nas tysiące, jeśli nie miliony razy mądrzejsze. Trudno ukryć fakt, że jeśli kiedykolwiek dojdzie do takiej kumulacji sztucznej inteligencji, z pewnością będzie ona miała ogromny wpływ na przyszłość ludzkości¹⁴.

Przewidywane ścieżki wiodące ku superinteligencji, wskazywane jak dotąd w literaturze przedmiotu, związane są z:

- (1) dalszym rozwojem silnej lub ogólnej sztucznej inteligencji;
- (2) transferem umysłu, czyli wyprodukowaniem inteligentnego programu poprzez zeskanowanie i ściśle odwzorowanie struktur obliczeniowych ludzkiego mózgu (czyli stworzenie kompletnej mapy sieci połączeń neuronalnych, zwanej konektomem);
- (3) poznaniem biologicznym, to znaczy usprawnieniem funkcjonowania mózgów biologicznych, na przykład poprzez manipulacje genetyczne i psychofarmakologię (leki, które poprawiają pamięć, koncentrację, mające dodawać energii intelektualnej);
- (4) bezpośrednim interfejsem mózg–komputer, zwłaszcza w postaci implan-tów, które miałyby umożliwić ludziom wykorzystanie silnych stron cyfrowych systemów obliczeniowych, perfekcyjnej pamięci, szybkiego i precyzyjnego wykonywania obliczeń arytmetycznych oraz szerokopasmowej transmisji da-

¹³ Por. T.J. Sejnowski, *Deep learning. Głęboka rewolucja. Kiedy sztuczna inteligencja spotka się z ludzką?*, tłum. P. Cypriański, Wydawnictwo Poltext, Warszawa 2019, s. 15-35.

¹⁴ Por. M. Ford, *Świt robotów. Czy sztuczna inteligencja pozbawi nas pracy?*, tłum. K. Łuniewska, Wydawnictwo CDP.pl, Warszawa 2017, s. 239-241.

nych – sprawiając, że powstałe systemy hybrydowe zdecydowanie prześcigną nierozszerzony mózg;

(5) stopniowym doskonaleniem sieci i organizacji łączących umysły pojedynczych osób ze sobą wzajemnie oraz z różnymi artefaktami i robotami. Koncepcja ta polega nie na podniesieniu potencjału intelektualnego jednostek w takim stopniu, aby każda z nich zyskała superinteligencję, lecz raczej na stworzeniu pewnego systemu złożonego z jednostek w taki sposób połączonych i zorganizowanych, aby mogły osiągnąć pewną formę superinteligencji¹⁵.

Prognozom przyszłości związanym z powyższymi przewidywanymi ścieżkami rozwoju towarzyszą oczywiście problemy etyczne oraz dylematy dotyczące granic postępu technologicznego, formułowane między innymi pod postacią pytania: jak oswoić sztuczną inteligencję?

Przykładowo Toby Walsh, autor książki *To żyje! Sztuczna inteligencja od logicznego fortepianu po zabójcze roboty*, przedstawia dziesięć promulgacji dotyczących tego, jak zmieni się ludzkie życie do 2050 roku. W dużym uproszczeniu odnoszą się one do następujących wydarzeń:

(1) „Mamy zakaz prowadzenia samochodów” (na drogach będą dominować samochody autonomiczne).

(2) „Codziennie będziemy z wizytą u lekarza” (lekarzem będzie zegarek fitness, smartfon, inne urządzenie komputerowe).

(3) „Marilyn Monroe wróci do filmu” (zaprogramowane awatary zmarłych aktorów będą grały w nowych, w pełni interakcyjnych filmach, zaś ludzie będą spędzać więcej czasu w światach, które nie istnieją, a w których wszyscy mogą być na przykład bogaci i sławni; nastąpi scalenie świata rzeczywistego, wirtualnego i rozszerzonego).

(4) „Komputer zatrudni i zwolni nas z pracy” (programy komputerowe będą decydować o zatrudnieniu ludzi, harmonogramie działań, wyrażać zgodę na urlop, monitorować i nagradzać).

(5) „Mówimy do mieszkania” (urządzenia w naszych domach pracują online; lodówki, tostery, bojler, światła, okna, doniczki, samochody są stale podłączone do sieci).

(6) „Robot napada na bank” (sztuczna inteligencja będzie stosowana nie tylko w systemach obronnych, ale i atakujących; nasili się problem cyberbezpieczeństwa).

(7) „Drużyna przegrywa z robotami” (ludzie-pilkarze przegrywają w piłkę nożną z zespołem złożonym z robotów).

(8) „Statki, samoloty i pociągi widma przemierzają świat” (automatyzacja zawodów spowoduje, że maszyny będą pozbawione ludzi-pilotów).

¹⁵ Por. N. B o s t r o m, *Superinteligencja. Scenariusz, strategie, zagrożenia*, tłum. D. Konow-rocka-Sawa, Wydawnictwo Helion, Gliwice 2016, s. 46-85.

(9) „Wiadomości telewizyjne będą powstawać bez udziału ludzi” (algorytmy zastąpią dziennikarzy, prezenterów, operatorów kamer).

(10) „Będziemy żyli po śmierci” (powszechne będzie tworzenie „cyfrowych sobowtórów”, które będą znały historie naszego życia i będą pocieszać naszą rodzinę po śmierci; będą się także pojawiać zamiast żyjących, na przykład w mediach społecznościowych, odpowiadać na e-maile, organizować spotkania i wydarzenia towarzyskie)¹⁶.

To jedynie hipotetyczne scenariusze przyszłości, ukazujące jednak, że skala potencjalnych problemów związanych z dalszym rozwojem technologii może być bardzo szeroka. Jak sztuczna inteligencja wpłynie na przykład na prawo i jego praktykę (czy komputer będzie mógł wyrażać zgody i zawierać umowy, posiadać własność, popełnić przestępstwo i być odpowiedzialnym za działania przestępcze)? Jak przełoży się na pracę ludzi (jakie zadania zautomatyzuje, jak wpłynie na pracowników fizycznych i umysłowych, czy istnieje alternatywa dla społeczeństwa opartego na pracy)? Jakie będzie przewidywane oddziaływanie na sprawiedliwość społeczną (kto skorzysta z rewolucji technologicznej, jak dystrybuować przyszłe aktywa bardziej sprawiedliwie)? Z powyższych pytań wynikają liczne kolejne, dotyczące możliwego przyszłego wpływu technologii na świat: Czy rozwój sztucznej inteligencji przyspiesza? Czy powstanie superinteligencji powinno być przedmiotem uzasadnionych obaw? Czy systemy sztucznej superinteligencji kiedykolwiek wymkną się spod kontroli? Jak minimalizować przyszłe zagrożenia? Jakie są korzyści i ryzyko związane z komputerami i robotami działającymi jak ludzie? Jak kolejne pokolenia, a przede wszystkim dzieci, będą odnosić się do sztucznej inteligencji? Czy kiedykolwiek człowiek będzie w stanie załadować sobie komputer?¹⁷ Co więcej, jeśli tak właśnie się stanie, to czy tradycyjne sposoby rozwoju i samodoskonalenia homo sapiens, czyli te związane z systemem edukacji, działalnością wychowawczą, szeroko rozumianą socjalizacją, medycyną konwencjonalną, rehabilitacją, terapią będą jeszcze potrzebne?¹⁸

¹⁶ Por. T. Walsh, *To żyje! Sztuczna inteligencja od logicznego fortepianu po zabójcze roboty*, tłum. W. Sikorski, Wydawnictwo Naukowe PWN, Warszawa 2018, s. 227-243.

¹⁷ Por. J. Kaplan, *Sztuczna inteligencja. Co każdy powinien wiedzieć*, Wydawnictwo PWN, Warszawa 2019, s. 113-191; J.J. Bryson, M.E. Diamantis, T.D. Grant, *Of, For, and By the People: The Legal Lacuna of Synthetic Persons*, „Artificial Intelligence and Law” 25(2017) nr 3, s. 273-291; *Legal Personhood: Animals, Artificial Intelligence and the Unborn*, red. V.A.J. Kurki, T. Pietrzykowski, Springer, Cham 2017.

¹⁸ Na przykład celem „Blue Brain Project” jest stworzenie wirtualnego symulatora ludzkiego mózgu, który będzie mówił i zachowywał się podobnie jak człowiek, z kolei projekt „Avatar 2045” zakłada zapewnienie cybernetycznej nieśmiertelności, czyli umożliwienie przetrwania świadomości po śmierci biologicznej ciała (zob. „Blue Brain Project,” EPFL, <https://www.epfl.ch/research/domains/bluebrain/>; 2045 Strategic Social Initiative, <http://2045.com/>).

Podobnych pytań można sformułować więcej. Potencjał zaawansowanej sztucznej inteligencji oraz strach przed skutkami ubocznymi jej rozwoju są coraz głośniej dyskutowane. Postęp w tej dziedzinie stanowi źródło refleksji nad wieloma zagadnieniami etycznymi i moralnymi. Jak się bowiem wydaje rozwój sztucznej inteligencji, a szczególnie podejmowane próby stworzenia superinteligencji, w sposób radykalny zakwestionują istotę człowieczeństwa i ontologicznej granicy jestestwa. Inspirują zatem do głębokiej rekonfiguracji antropologicznych założeń na temat człowieka¹⁹.

Sztuczna inteligencja rodzi wyzwania także dla doktryn filozoficznych oraz religijnych. Intelligentne maszyny mogą rzucić obiektywne światło na fundamentalne kwestie dotyczące natury ludzkich umysłów, istnienia wolnej woli oraz tego, czy o niebiologicznych podmiotach można twierdzić, że są żywe. Te intelektualne pytania łączą się z obawami, że sztuczna inteligencja może zagrozić źródłom utrzymania, a nawet życiu ludzi. Owe obawy – co warto również dodać – podsycane są przez powracający w literaturze i filmach motyw „buntu robotów”, datowany przynajmniej od napisanej w 1920 roku sztuki *R.U.R.*²⁰ czeskiego dramaturga Karela Čapka, któremu przypisuje się stworzenie określenia „robot” od czeskiego słowa „robota”, oznaczającego przymusowo wykonywaną pracę²¹.

W świetle ukazanych w artykule informacji zastanowić się można nad hipotetycznymi scenariuszami przyszłości. Staje się ona coraz częściej obiektem zainteresowania badaczy społecznych, którzy starają się konstruować jej wizje, wskazując jednocześnie na warunki i możliwości kształtowania ich przez ludzi. Ich twórcy często szukają „kamienia filozoficznego”, przy pomocy którego dałoby się współczesny świat zrozumieć i ulepszyć. Jedni patrzą w przyszłość z punktu widzenia niebezpieczeństw zagrażających ludzkości, chcąc zrozumieć ich przyczyny, pragnąc ostrzec. Inni w swoich wizjach przejmują się nierozwiązanymi problemami i poszukują alternatyw rozwojowych oraz rekomendacji dla działań praktycznych. Jeszcze inni spoglądają w przyszłość właśnie z punktu widzenia sił napędowych cywilizacji, takich jak nauka i technika²².

¹⁹ Por. M. Lipowicz, *Człowieczeństwo jako (nie)zbędna kategoria refleksji pedagogicznej? O ponowoczesnym kryzysie teorii wychowania w obliczu wyzwania trans- i posthumanizmu*, „Studia z Teorii Wychowania” 8(2017) nr 2(19), s. 35-57.

²⁰ Tytuł ten jest skrótem wyrażenia „Uniwersalne Roboty Rossuma” (zob. K. Čapek, *R.U.R.*, tłum. M.M. Lemańczyk, Wydawnictwo CM, Warszawa 2023).

²¹ Por. K. a p l a n, dz. cyt., s. 89n.

²² Por. L.W. Z a c h e r [Polska Akademia Nauk Komitet Prognoz „Polska 2000 Plus”], *Gry o przyszłe światy*, Warszawska Drukarnia Naukowa PAN, Warszawa 2006, s. 17n.

PRZYSZŁOŚĆ: RAJSKA, PIEKIELNA, TRANSCENDENTALNA?

Prognozy konstruowane w odniesieniu do sztucznej inteligencji bywają zróżnicowane. Od optymistycznych, utopijnych, po „czarne”, dystopijne, „w których technologiczne szaleństwo w połączeniu z nieokiełznaną biologiczną szarlatanerią doprowadza do zapowiadanej po wielokroć klęski człowieka i człowieczeństwa”²³.

W tej części artykułu można przybliżyć trzy scenariusze przyszłości i wynikające z nich trzy sposoby oceny obecnego oraz przyszłego stanu człowieka i człowieczeństwa, które układają się w triadę: Raj – Piekło – Triumf. Ich autorem jest Joel Garreau, autor książki *Radykalna ewolucja. Czy człowiek udoskonalony przez naukę i technikę będzie jeszcze człowiekiem?*²⁴. Scenariusze te stanowią interesujące metafory, oddające istotę rozwoju technologicznego.

Przybliżając pierwszy z nich – rajski – powiązać go należy z zagorzałym optymizmem, że już niedługo ludzi czeka coś, co nieodparcie kojarzy się z chrześcijańską wizją raju. Zakłada on między innymi, że człowiek przekroczy granice swojej natury albo dzięki komputerom, albo poprzez inżynierię genetyczną, umożliwiającą zarówno przeprogramowanie organizmu, usunięcie niedoskonałości swojego ciała, jak i dotrzymanie przez ewolucję biologiczną kroku gwałtownemu rozwojowi systemów komputerowych. W tym scenariuszu w ciągu najbliższych dekad radykalnie spowolniony zostanie proces starzenia się, znacznej części chorób uda się zapobiegać, łatwiej będzie też naprawić ich skutki. Większość ludzi wspomagana będzie na co dzień przez urządzenia pierwotnie przeznaczone dla osób z niepełnosprawnościami. Implanty nerwowe pozwolą członkom rodziny cieszyć się swoją obecnością, pomimo dzielącej ich fizycznej odległości. Liczba ludności na świecie ukształtuje się na poziomie dwunastu miliardów. Gospodarka światowa będzie w rozkwicie, stąd zaspokojone zostaną potrzeby większości ludzi w zakresie pożywienia, dachu nad głową i bezpieczeństwa. Co kluczowe, coraz silniej zacierać się będzie granica między światem ludzi i światem maszyn. Inteligencja maszynowa wzorowana będzie na ludzkiej, a ludzka będzie wspomagana przez maszyny. Dzięki obwodom komputerowym wszczepianym bezpośrednio do mózgu ludzie będą dysponować większą niż kiedykolwiek zdolnością logicznego myślenia, zapamiętywania i percepcji. Z kolei osobowy charakter, umiejętności i wiedza wielu maszyn będą się wywodzić z ludzkiego umysłu. W obydwu przypadkach to implanty umożliwią zintegrowanie mózgu z komputerem. Całkowita moc obliczeniowa dostępna dla ludzkości, na którą złożą się mózgi wszystkich

²³ P. Z a w o j s k i, *Technokultura i jej manifestacje artystyczne. Medialny świat hybryd i hybrydyzacji*, Wydawnictwo Uniwersytetu Śląskiego, Katowice 2016, s. 24.

²⁴ Por. J. G a r r e a u, dz. cyt., s. 91-138.

ludzi i cała wytworzona sztuczna inteligencja w dziewięćdziesięciu dziewięciu procentach, realizowana będzie przez maszyny. Do tego czasu uda się także zbadać większość obszarów ludzkiego mózgu, a wiedzę tę wykorzysta się przy tworzeniu komputerów najnowszej generacji. Sieci neuronowe wzorowane na aktywności prawdziwych neuronów będą działać znacznie szybciej, a ich zdolność przetwarzania i zapamiętywania danych przewyższy możliwości mózgu homo sapiens. Ludzie będą mieli w oczach urządzenia – wszczepione na stałe lub zakładane niczym soczewki kontaktowe – pozwalające na oglądanie w polu widzenia trójwymiarowych obrazów wygenerowanych przez komputer. Do interaktywnej komunikacji z ogólnosiwiatową siecią używane będą też implanty w uszach, które jednocześnie wzmocnią słuch. Z kolei implanty w mózgu pozwolą na bezpośrednią wymianę myśli z komputerem. W razie potrzeby będzie można dokupić sobie więcej pamięci długotrwałej bądź też dodatkowy moduł myślenia logicznego. W ten sposób homo sapiens stworzy rzesze własnych następców poprzez przyspieszoną ewolucję. W rajskiej wizji przyszłości człowiek przestaje być człowiekiem w dzisiejszym rozumieniu tego słowa, to jest nieuchronne przeznaczenie potomnych²⁵.

Przykładem takiego technologicznego optymizmu jest wizja przyszłości zawarta w dokumencie *Converging Technologies for Improved Human Performance: Nanotechnology, Biotechnology, Information Technology and Cognitive Science* (Współdziałanie innowacji naukowo-technicznych w zakresie zwiększania możliwości człowieka: nanotechnologia, biotechnologia, technologie informacyjne i kognitywistyka), którego redaktorami są Mihail C. Roco i William Sims Bainbridge z National Science Foundation²⁶. Raport ten zakłada coraz większą integrację człowieka i maszyny oraz wskazuje na korzyści („Raj”), jakie taka radykalna ewolucja ma przynieść, między innymi:

(a) ciało ludzkie ma stać się bardziej wytrzymałe, zdrowe i pełne energii, łatwiejsze w leczeniu oraz odporne na różnorakie stresy, zagrożenia biologiczne i procesy starzenia; każdy w dowolnym miejscu na świecie będzie mógł uzyskać natychmiast wszelkie potrzebne mu informacje;

(b) bezpośrednie połączenia między ludzkim mózgiem a maszyną mają zmienić charakter pracy w przemyśle, sposób kierowania samochodami, zapewnić przewagę militarną, przyczynić się do powstania nowych dyscyplin sportowych i środków wyrazu artystycznego, a także trybów kontaktów międzyludzkich;

²⁵ Por. tamże.

²⁶ Zob. *Converging Technologies for Improved Human Performance: Nanotechnology, Biotechnology, Information Technology and Cognitive Science*, red. M.C. Roco, W.S. Bainbridge, Kluwer Academic Publishers, New York 2003.

(c) noszone przy sobie czujniki będą monitorować nie tylko stan zdrowia właściciela, ale i parametry środowiska, ewentualne skażenia chemiczne, inne typy zagrożeń oraz w razie potrzeby dostarczą danych o lokalnych firmach, zasobach naturalnych i tak dalej;

(d) poziom bezpieczeństwa narodowego istotnie się zwiększy dzięki wysoce skomputeryzowanym typom broni, bezzałogowym maszynom bojowym, „inteligentnym” materiałom zmieniającym swoje właściwości w zależności od potrzeb, superefektywnym systemom rozpoznawania oraz skutecznym sposobom neutralizacji skutków użycia broni biologicznej, chemicznej, radiologicznej i atomowej;

(e) ziszcza się także nadzieje pokładane w podboju kosmosu dzięki rakietom nośnym nowej konstrukcji, bazom pozaziemskim budowanym przez roboty oraz intratnej eksploatacji surowców naturalnych na Księżycu, Marsie i pobliskich planetoidach²⁷.

Scenariusz rajski zakłada zatem, że pod koniec dwudziestego pierwszego wieku osiągnięty zostanie trwały światowy pokój i powszechny dobrobyt. Dzięki sztucznej inteligencji nastąpi rozbudowa sieci komunikowania się między ludźmi, a ludzkość – być może – stanie się jednym wielkim rozproszonym „mózgiem”. Drugi ze scenariuszy – piekielny – jest zwierciadlanym odbiciem scenariusza niebiańskiego. Przewiduje, że wysoki stopień rozwoju technicznego będzie stanowił zagrożenie dla ludzkości. Pojawią się nowe, potężne środki do czynienia potwornego zła, mogące zostać wykorzystane przez rozmaitych ekstremistów. Rozważając perspektywy stwarzane przez genetykę, robotykę, informatykę i nanotechnologię²⁸, kreśli się tutaj wizję wywołania „białej dżumy”, zabijającej bardzo wiele, lecz wybranych ofiar, przykładowo ludzi określonej rasy. Dałoby to możliwość pokierowania naszą ewolucją w ten sposób, że ludzkość rozpadłaby się na szereg odrębnych gatunków o różnym stopniu rozwoju, co podważyłoby ideę równości, stanowiącą kamień węgielny demokracji. Kolejna wizja w scenariuszu piekielnym zakłada, że superinteligentne roboty sprowadziłyby życie swoich twórców do poziomu wegetacji, to znaczy realny stałby się scenariusz „szarego szlamu”, „szarej mazi” (ang. „grey goo”²⁹). Termin ten został wprowadzony przez pioniera nanotechnologii, fizyka Kima Erica Drexlera, autora książki *Engines of Creation* („Motory tworzenia”). Oznacza on hipotetyczny punkt apokaliptycznego scenariusza rozwoju ludzkości, w którym samoreplikujące się nanomechanizmy wyrrywają się spod kontroli i przekształcają całą biosferę w swoje

²⁷ Tamże.

²⁸ Technologie dwudziestego pierwszego wieku są oznaczone kryptonimem GNR (genetyka, nanotechnologia, robotyka) bądź GRIN (genetyka, robotyka, informatyka i nanotechnologia).

²⁹ K.E. Drexler, *Engines of Creation: The Coming Era of Nanotechnology*, Anchor Books, Doubleday 1986, s. 127, 208.

kopie, zabijając wszystko, co żyje na ziemi. Sama nazwa „szara maź” ma sugerować bezpostaciową, rozprzestrzeniającą się masę reprodukowaną przez maszyny. Zakłada się zatem zagładę świata wywołaną przez samoreplikujące się, niezniszczalne nanomechanizmy, wysysające wszelkie substancje życiowe z żywych organizmów. W przeciwieństwie do broni atomowej, stworzone genetycznie patogeny, superinteligentne roboty, małeńkie asemblery i wirusy komputerowe, raz wypuszczone, zostaną powielone tryliony razy i trudno je będzie powstrzymać. Istota scenariusza piekielnego to zagłada ludzkości oraz zatracenie człowieczeństwa, a jedyną realistyczną alternatywą, aby ta wizja przyszłości nie spełniła się jest „samoograniczenie się”, to znaczy odrzucenie tych rozwiązań technicznych, które są nazbyt niebezpieczne, i niepodejmowanie dalszych badań w niektórych obszarach wiedzy³⁰.

Jak natomiast wskazuje trzeci scenariusz – triumfalny – nie powinniśmy jako ludzie odrzucać zdobyczy nowych technologii, ale musimy mieć głęboką świadomość drżającego w nich zagrożenia. Przykładowo w maju 2014 roku w „The Independent” ukazał się artykuł, którego współautorem był słynny fizyk Stephen Hawking z Uniwersytetu w Cambridge. W tekście tym przestrzegano przed zagrożeniami związanymi z szybkim rozwojem sztucznej inteligencji. Poza Hawkingiem podpisali się pod nim Max Tegmark, laureat Nagrody Nobla Frank Wilczek oraz informatyk Stuart Russell z Uniwersytetu Kalifornijskiego. Zdaniem autorów tego artykułu stworzenie superinteligencji, czyli myślącej maszyny, może być największym wydarzeniem w historii ludzkości. Naukowcy zasugerowali także, że taki superkomputer mógłby przechytrzyć rynki finansowe, przewyższyć ludzi inteligencją, przejąć nad nami kontrolę, a nawet stworzyć broń, której nie jesteśmy w stanie zrozumieć. Przy okazji uczeni ci przestrzegali, że uznanie ich konstatacji za stwierdzenie typu *science-fiction* może okazać się największym błędem w historii³¹. Żarliwej obronie podmiotowości ludzkiej w czasach rozwoju sztucznej inteligencji powinno zatem towarzyszyć przekonanie o tym, że świadome używanie technologii w dużej mierze winno wiązać się jednocześnie z częściowym ich odrzuceniem. Kluczowy w tej prognozie jest apel, aby ludzie wykorzystywali technikę i technologię przede wszystkim do nawiązywania więzi między sobą i to na szeroką skalę. Dzięki sztucznej inteligencji będzie można bowiem two-

³⁰ Por. Garreau, dz. cyt., s. 139-193.

³¹ Zob. S. Hawking, S. Russell, M. Tegmark, F. Wilczek, *Stephen Hawking: Transcendence Looks at the Implications of Artificial Intelligence – But Are We Taking AI Seriously Enough?*, „The Independent” May 1, 2014, <https://www.independent.co.uk/news/science/stephen-hawking-transcendence-looks-implications-artificial-intelligence-are-we-taking-ai-seriously-enough-9313474.html>.

rzyć nowe, alternatywne rzeczywistości³². W związku z tym możliwe będzie też wnoszenie się ku „transcendencji” (transcendowanie, czyli przekraczanie granic czasowych, przestrzennych, psychologicznych i tym podobnych; przekraczanie skończoności ku horyzontowi nieskończoności – rzeczywistości absolutnej). Sednem scenariusza triumfalnego jest poszukiwanie złożonej, ewoluującej i innowacyjnej transcendencji, która jednak nigdy nie doprowadzi do superinteligencji, w przeciwieństwie do scenariusza rajskiego i piekielnego. Kluczową miarą sukcesu jest w tej wizji wspomniane zintensyfikowanie więzi między ludźmi, a nie połączeń między tranzystorami. Transcendencja w scenariuszu triumfalnym jest społeczna, a nie jednostkowa. Jej miarą jest to, do jakiego stopnia ludzie będą razem się zmieniać. Dzięki temu charakteryzowany scenariusz ma nieskończenie wiele wersji, choć opierają się one na wspólnych zasadach:

(a) dzieje człowieka to przecieranie niemożliwych ścieżek ku mało prawdopodobnym przyszłościom na przekór oczywistym i nieuniknionym siłom historii;

(b) źródło tych zmagających leży w zdolności do stawiania do nierównej walki, gdy wymaga tego sytuacja;

(c) nawet jeśli rozwój techniki przebiega wzdłuż krzywej wykładniczej, nie oznacza to, że ludzie nie mogą twórczo i w sposób nieprzewidywany kształtować wpływu techniki na ludzką naturę i społeczeństwo³³.

Jak podkreśla się w trzeciej wizji, gatunek ludzki ma potencjał do przemiany w coś wykraczającego daleko poza dzisiejsze rozumienie ludzkiej natury. Wizja załadowania części mózgu do komputera prowadzi donikąd. Zamiast tego należy dostrzec różne warianty transcendencji. Scenariusz transcendentny jest rozwinięciem i punktem odniesienia dla scenariusza triumfalnego, w którym to nie technika panuje nad człowiekiem, tylko człowiek nad techniką. Teza o transcendencji opiera się na trzech przesłankach:

„1. Niezaprzeczalna przewaga konkurencyjna zapewniana przez genetykę, robotykę, informatykę i nanotechnologię osobom przyjmującym ich zdobycze z przyczyn ekonomicznych, medycznych, edukacyjnych, wojskowych albo artystycznych pozwala przypuszczać, że będą się one nadal rozwijać w narażającym tempie.

³² Na przykład metaverse – wirtualny świat naśladujący rzeczywistość, w którym ludzie jako awatary wchodzą ze sobą w interakcje w trójwymiarowej przestrzeni. Jego istotą jest zanurzenie się w świecie 3D, czyli tak zwana immersyjność i poczucie jego realizmu z punktu widzenia awatara. Interfejs między rzeczywistością a metaverse zapewniają gogle VR i rękawice, dzięki którym możliwa jest manipulacja wirtualnymi obiektami (zob. M. B a l l, *Metawersum. Jak internet przyszłości zrewolucjonizuje świat i biznes*, tłum. K. Mironowicz, Wydawnictwo MT Biznes, Warszawa 2022).

³³ Por. G a r r e a u, dz. cyt., s. 194-230.

2. Wiele z tych technologii – w tym «projektowane dzieci», ulepszone poznanie, poprawki metaboliczne, medycyna opóźniająca starzenie i wiele innych – może przyczynić się do zmiany kondycji człowieka jako takiego. Skoro modyfikacjom poddaje się ludzki mózg, wspomnienia, metabolizm, potomstwo i osobowość, logiczne wydaje się, że tego rodzaju procedury wpłyną prawdopodobnie na to, co oznacza być człowiekiem.

3. Dzieje podobnie rewolucyjnych technologii sugerują, że należy spodziewać się niezamierzonych konsekwencji. Ich rezultaty niejednokrotnie okazały się zaskakujące³⁴.

Jeżeli uznać powyższe za racjonalne, wówczas otrzymamy wielce prawdopodobny scenariusz przyszłości o dużej sile oddziaływania. A zatem to człowiek panuje nad techniką. W nawiązaniu zaś do powyższego stwierdzenia, że wizja załadowania części mózgu do komputera prowadzi donikąd, można zacytować Stephena Wolframa, pioniera rozwijania i wykorzystywania myślenia komputacyjnego (ang. computational thinking), rozumianego jako proces znajdowania rozwiązań problemów z różnych dziedzin przy świadomym wykorzystaniu metod i narzędzi informatycznych. „Założmy, że dojdzie do tego, że ludzką świadomość będzie się dało zamienić na postać cyfrową, w pełni zwirtualizowaną i tak dalej. Wkrótce będziemy mieli skrzynkę z bilionem dusz. Bilion dusz w skrzynce, wszystkie zwirtualizowane. W skrzynce będzie zachodziła komputacja molekularna – może bazująca na biologii, może nie. Ale skrzynka będzie w stanie wykonywać wszelkiego rodzaju skomplikowane procesy. Założmy, że obok skrzynki leży kamień. W kamieniu od zawsze dokonują się wszelkiego rodzaju skomplikowane procesy, gdy wszystkie rodzaje cząstek subatomowych wykonują najróżniejszego rodzaju działania. Jaka jest różnica między kamieniem a skrzynką z bilionem dusz? Odpowiedź brzmi: szczegóły tego, co dzieje się w skrzynce, bazują na długiej historii cywilizacji, włączając w to treści, jakie ludzie oglądali na YouTube poprzedniego dnia, natomiast skała ma długą historię geologiczną, ale nie jest to konkretna historia naszej cywilizacji. Uświadomienie sobie, że nie istnieje realna różnica między inteligencją a zwykłą komputacją, prowadzi nas do wizji takiej przyszłości – zwieńczenia naszej cywilizacji w postaci skrzynki z bilionem dusz, z których każda zasadniczo gra w grę wideo, na wieczność. Jaki to będzie miało «cel»?»³⁵.

Podsumowując tę część artykułu, w optyce „rajskiej” ludzkość ma zostać diametralnie udoskonalona dzięki postępowi w zakresie sztucznej inteligencji. W wizji „piekielnej” przewiduje się katastroficzną zagładę, oznaczającą kres człowieka w znanej dziś postaci. Wydaje się, że każdy z tych scenariuszy ma

³⁴ Tamże, s. 238.

³⁵ S. W o l f r a m, *Sztuczna inteligencja i przyszłość cywilizacji*, w: *Człowiek na rozdrożu*, s. 303n.

szansę się zrealizować. Z kolei trzeci scenariusz „triumfalny” jest bardziej skomplikowany, niejednoznacznie zdeterminowany. Według tej wizji to nadal ludzie, a nie AI czy AGI, będą wybierać kształt własnej przyszłości, kierując się wyznawanymi wartościami.

JAK OSWOIĆ SZTUCZNĄ INTELIGENCJĘ?

W nawiązaniu do ukazanych prognoz łączą się dwie, całkowicie odmienne wizje przyszłych wydarzeń – dobre AI kontra złe AI; połączenie Paruzji i Apokalipsy³⁶.

Z jednej strony, przyszłość jawi się jako czas pomyślności i przyjemności oraz świat bezpośredniej natychmiastowej łączności, nieograniczonej wiedzy i nieprawdopodobnych udogodnień. Moce obliczeniowe komputerów będą narastały, narodzą się innowacyjne, prężne gałęzie gospodarki i nowe rodzaje zajęć – cyberzawody. Osiągnięcie przez roboty świadomości ma, w znacznej mierze, przyczynić się do wzrostu ich użyteczności w społeczeństwie. Mogłyby na przykład podejmować samodzielne decyzje, pracować jako sekretarki, lokaje, asystenci i pomocnicy³⁷.

Z drugiej strony, obecne wektory rewolucji technologicznej mogą prowadzić do fazy przejścia przez sztuczną inteligencję kontroli nad kluczowymi funkcjami poznawczymi człowieka. W świetle tej hipotezy nasuwa się myśl, że jeden tylko rodzaj globalizacji będzie naprawdę wszechobejmujący: cyfryzacja „wszystkiego”, jakaś forma totalitaryzmu. Wtedy jedna globalna świadomość zatroszczyłaby się o wszystko. Dziś nie sposób określić jeszcze, do czego to „wszystko” będzie się odnosić³⁸, ale taka wizja sztucznej inteligencji, jak już podkreślono w artykule, pojawia się często. Obawy budzi scenariusz, w którym AI przejmuje kontrolę oraz przewiduje zamiary człowieka i potrafi ustalić sposób realizacji określonego postępowania w ich obliczu.

Istnieje zatem mroczna prognoza przyszłości, w której sztuczna inteligencja jest postrzegana jako zagrożenie dla ludzkiej egzystencji i przyczynia się do powstania koszmarnego świata. Niewykluczone, że do połowy tego wieku komputery staną się tak sprawne, że będą mogły zarządzać dużymi miastami, a nawet państwami. Można im będzie powierzyć stałą kontrolę nad energetyką, przepływem surowców i produktów w obrocie gospodarczym, bankowością, handlem, transportem publicznym, zaopatrzeniem w wodę i usuwaniem nie-

³⁶ Por. B r o c k m a n, dz. cyt., s. 9.

³⁷ Por. M. K a k u, *Wizje czyli jak nauka zmienia świat w XXI wieku*, tłum. K. Pesz, Wydawnictwo Prószyński i S-ka, Warszawa 2000, s. 147-197.

³⁸ Por. A. Z y b e r t o w i c z i n., *Samobójstwo Oświecenia? Jak neuronauka i nowe technologie pustoszą ludzki świat*, Wydawnictwo Kasper, Kraków 2015, s. 430-440.

czystości, środowiskiem naturalnym. Każdy defekt w obwodach sterujących ich systemem, awaria, pętla sprzężenia zwrotnego będą grozić upadkiem lub paraliżem cywilizacji o tragicznych dla ludzkości skutkach („szalone roboty”, „roboty-zabójcy”). Niewykluczone również, że systemy sztucznej inteligencji rozrosną się, podobnie jak biurokracja, do ogromnych rozmiarów. W tym sensie nieumyślnym zagrożeniem mogą stać się posłuszne, „spolegliwe roboty”, które będą znakomicie spełniać swoją misję³⁹. Co więcej, jeśli któraś ze ścieżek powstania superinteligencji się ziści, to pojawić się może „stwór” wyposażony w nadludzką inteligencję, który mógłby zagrozić przyszłości gatunku ludzkiego, chociaż mielibyśmy jako ludzie jedną przewagę – to my bylibyśmy jego twórcami⁴⁰.

Zasygnalizowane potencjalne zagrożenia wskazują, że sztuczna inteligencja będzie musiała podlegać coraz ściślejszej kontroli, aby nie pojawiły się niepożądane konsekwencje. Roboty należy wyposażać w urządzenia zabezpieczające, które nie pozwoliłyby im przejąć władzy na Ziemi. Być może powinna powstać nowa gałąź sztucznej inteligencji zajmująca się strictly problemem utrzymywania układów sztucznej inteligencji pod kontrolą⁴¹. W praktyce zagadnienie kontroli nie jest jednak proste. Wręcz przeciwnie – jawi się jako mocno skomplikowane, a sama technologia wydaje się być współczesnym paradoksem, gdyż z jednej strony tworzy ją człowiek, ale z drugiej tak naprawdę nic o niej nie wie⁴².

Należy podkreślić, że ostrzeżenia dotyczące niekontrolowanego postępu technologicznego są znane od lat. Nie są one natomiast szerzej publikowane, a dotychczasowe oficjalne dyskusje wydają się niedostateczne. Dzieje się tak między innymi dlatego, że – jak pisze w swoim esej *Dlaczego przyszłość nas niepotrzebuje?* Bill Joy⁴³ – opisywanie zagrożeń nie sprzyja osiąganiu zysków, a najnowocześniejsze technologie mają wyraźnie komercyjne zastosowanie i rozwija się je niemal wyłącznie w korporacjach. W wieku triumfującego komercjalizmu wraz z nauką dostarczają serii innowacji fenomenalnie lukratywnych.

Rozważając o przyszłości inteligentnych maszyn i o tym, czy przejmą one władzę i zaczną same decydować o sobie, warto dodać, że najpierw ktoś lub

³⁹ Por. K a k u, dz. cyt., s. 174-197.

⁴⁰ Por. B o s t r o m, dz. cyt., s. 11.

⁴¹ Por. K a k u, dz. cyt., s. 174-197.

⁴² Interesujące scenariusze w kontekście rozwoju technologii prezentuje Natalia Hatałska (por. N. H a t a ł s k a, *Wiek paradoksów. Czy technologia nas ocali?*, Wydawnictwo Znak, Kraków 2021). Każdy ze scenariuszy można zawsze zmienić, ostatecznie przyszłość zależy od naszych wyborów tu i teraz.

⁴³ Por. B. J o y, *Dlaczego przyszłość nas nie potrzebuje?*, w: *Wybierz czerwoną pigułkę: Nauka, filozofia i religia w Matrix*, red. G. Yeffeth, tłum. W. Derechowski, Wydawnictwo Helion, Gliwice 2003, s. 229-245.

coś musi zdefiniować zadanie maszyny, czyli to, co maszyna ma wykonać. Dla człowieka określanie celów jest zazwyczaj powiązane z warunkami biologicznymi i psychologicznymi, historią osobistą, środowiskiem kulturowym, historią cywilizacji. Zamiary są unikalną cechą ludzi. Jeśli chodzi o maszyny, to ludzie nadają im cel po ich skonstruowaniu⁴⁴. Stąd, rozważając przyszłość, nie należy pomijać problematyki związanej z zadaniami AI w oparciu o prawa matematyki i fizyki. Inteligentnym systemom można wgrać szkodliwe cele, dlatego też kluczowe będzie zaprojektowanie technologicznej infrastruktury zdolnej do wykrywania i kontrolowania zachowań systemów szkodliwych. Konieczne jest tworzenie prawnych i ekonomicznych ram stymulujących zachowania AI na bazie intelektualnych i technologicznych zasobów, jakimi dysponujemy jako ludzie⁴⁵.

Jednak nie po raz pierwszy jesteśmy w sytuacji, kiedy nowo powstała technologia wydaje się stwarzać zagrożenie dla naszej egzystencji. Wynalezienie bomby atomowej oraz tworzenie jądrowych arsenałów groziło i stale grozi zagładą świata. Z kolei kiedy po raz pierwszy pojawiła się technologia rekombinowania DNA, towarzyszył temu lęk, że zmodyfikowane genetycznie organizmy wydostaną się na wolność i doprowadzą do śmierci wielu ludzi na całym świecie. Jak się wydaje, osiągnięcia w dziedzinie uczenia maszynowego stwarzają stosunkowo niewielkie zagrożenia w porównaniu z bronią nuklearną czy śmiertelnościami organizmami. Stąd można dywagować, że do sztucznej inteligencji również się przyzwyczaimy, co zresztą już się dzieje⁴⁶.

Pytanie: jak oswoić sztuczną inteligencję? – pozostaje otwarte. Wiąże się z nim bowiem kolejne: czy ludzie będą w stanie kontrolować w pełni zrealizowaną i samodzielnie ulepszającą się sztuczną inteligencję? Dziś jeszcze nie wiadomo, który ze scenariuszy najpełniej oddaje istotę nadchodzącej przyszłości – „Raj”, „Piekło” czy „Triumf”. Sztuczna inteligencja z jednej strony przynosi niebywałe obietnice „innego”, „lepszego” życia, ale z drugiej strony, stanowi tak wielki potencjał, że jest zarazem ogromnym zagrożeniem. Kiedy roboty będą stały się coraz inteligentniejsze i coraz bardziej podobne do ludzi, niebezpieczeństwo zagrażające z ich strony rzeczywiście może przyjąć realniejszą postać. Jedno jest natomiast pewne, nauka o sztucznej inteligencji rozwija się, w języku technicznym AI ewoluuje. Dlatego przyszłość człowieka jest związana z rozwojem sztucznej inteligencji.

⁴⁴ Por. W o l f r a m, dz. cyt., s. 289n.

⁴⁵ Por. S. O m o h u n d r o, *Punkt zwrotny w dziedzinie sztucznej inteligencji*, w: *A ty co sądzisz o myślących maszynach? Wzrywy przyszłości wybitnych umysłów ery sztucznej inteligencji*, red. J. Brockman, tłum. K. Kubala, Wydawnictwo Naukowe PWN, Warszawa 2020, s. 31.

⁴⁶ Por. S e j n o w s k i, dz. cyt., s. 40.

*

Sztuczna inteligencja to narracja współczesności. Ukazanie wszystkich kwestii związanych z tym zagadnieniem, wyjaśnienie niejasności, ocena zarówno tychże narracji, jak i samego zjawiska, a także dokonanie ewaluacji wizji przyszłości, które są przedstawiane przez zwolenników i oponentów, nie jest możliwe w jednej publikacji. Zakres analiz jest w tym obszarze problemowym bardzo szeroki, zaś tematykę wpływu technologii na przykład na edukację, pracę, medycynę, kulturę i tym podobne, można rozpatrywać także odrębnie. Dyskurs na temat sztucznej inteligencji wydaje się tak naprawdę dopiero rozpoczynać. Problematyka ta jest złożona, a w jej ramach pojawia się coraz więcej pytań niż odpowiedzi. Postęp w tej dziedzinie i zachodzące wraz z nim zmiany będą natomiast fundamentalne dla przyszłości ludzi i dla świata. Wielu naukowców uważa, że podążamy w nowe stulecie bez planu, sterowania i hamulców: „Czy jesteśmy za daleko, żeby zmienić drogę? Być może nie, ale nawet nie próbujemy tego uczynić, a ostatnia szansa, by zacząć kierować – ten punkt, skąd można się cofnąć – zbliża się szybko”⁴⁷. „Jeszcze nie jest jasne, czy wygramy, czy przegramy, przetrwamy, czy zginiemy przez te technologie”⁴⁸.

BIBLIOGRAFIA / BIBLIOGRAPHY

- Adcock, Steve. *These Are the Exact Skills You Need to Get a Job in Artificial Intelligence*. Ladders, Ferbruary 5, 2020. <https://www.theladders.com/career-advice/these-are-the-exact-skills-you-need-to-get-a-job-in-artificial-intelligence>.
- Ball, Matthew. *Metawersum: Jak internet przyszłości zrewolucjonizuje świat i biznes*. Translated by Katarzyna Mironowicz. Warszawa: Wydawnictwo MT Biznes, 2022.
- Bostrom, Nick. *Superinteligencja: Scenariusze, strategie, zagrożenia*. Translated by Dorota Konowrocka-Sawa. Gliwice: Wydawnictwo Helion, 2016.
- Brockman, John. “Wprowadzenie: O nadziejach i pułapkach związanych z AI.” In *Człowiek na rozdrożu: Sztuczna inteligencja – 25 punktów widzenia*. Edited by John Brockman. Translated by Marcin Machnik. Gliwice: Wydawnictwo Helion, 2020.
- Bryson, Joanna J., Mihailis E. Diamantis, and Thomas D. Grant. “Of, For, and By the People: The Legal Lacuna of Synthetic Persons.” *Artificial Intelligence and Law*, no. 25 (3) (2017): 273–91.
- Čapek, Karel. *R.U.R.* Translated by Martyna M. Lemańczyk, Wydawnictwo CM, Warszawa 2023.

⁴⁷ J o y, dz. cyt., s. 245.

⁴⁸ Tamże, s. 251.

- Communication from the Commission to the European Parliament, the European Council, the Council. *The European Economic and Social Committee and the Committee of the Regions on Artificial Intelligence for Europe*, Brussels, 25.04.2018, COM(2018) 237 final. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN>.
- Drexler, Kim Eric. *Engines of Creation: The Coming Era of Nanotechnology*. Doubleday: Anchor Books, 1986.
- European Commission. *White Paper: On Artificial Intelligence; A European Approach to Excellence and Trust*, Brussels, 19.02.2020, COM(2020) 65 final. https://commission.europa.eu/system/files/2020-02/commission-white-paper-artificial-intelligence-feb2020_en.pdf.
- Flasiński, Mariusz. *Wstęp do sztucznej inteligencji*. Warszawa: Wydawnictwo PWN, 2018.
- Ford, Martin. *Świt robotów: Czy sztuczna inteligencja pozbawi nas pracy?* Translated by Katarzyna Luniewska. Warszawa: Wydawnictwo CDP.pl, 2017.
- Furmanek, Waldemar. "Człowiek jako obiekt badań humanistycznej pedagogiki pracy." *Labor et Educatio*, no. 2 (2014): 9–30.
- Garreau, Joel. *Radykalna ewolucja: Czy człowiek udoskonalony przez naukę i technikę będzie jeszcze człowiekiem?* Translated by Agnieszka Kloch and Aleksander Michalski. Poznań: Wydawnictwo Prószyński i S-ka, 2005.
- Hatańska, Natalia. *Wiek paradoksów: Czy technologia nas ocali?* Kraków: Wydawnictwo Znak, 2021.
- Hawking, Stephen, Stuart Russell, Max Tegmark, and Frank Wilczek. "Stephen Hawking: Transcendence Looks at the Implications of Artificial Intelligence – But Are We Taking AI Seriously Enough?" *The Independent*. May 01, 2014. <https://www.independent.co.uk/news/science/stephen-hawking-transcendence-looks-implications-artificial-intelligence-are-we-taking-ai-seriously-enough-9313474.html>.
- Joy, Bill. "Dlaczego przyszłość nas nie potrzebuje?" In *Wybierz czerwoną pigułkę: Nauka, filozofia i religia w Matrix*. Edited by Glenn Yeffeth. Translated by Wojciech Derechowski. Gliwice: Wydawnictwo Helion, 2003.
- Kaku, Michio. *Wizje czyli jak nauka zmieni świat w XXI wieku*. Translated by Karol Pesz. Warszawa: Wydawnictwo Prószyński i S-ka, 2000.
- Kaplan, Jerry. *Sztuczna inteligencja: Co każdy powinien wiedzieć*. Warszawa: Wydawnictwo PWN, 2019.
- Krause, Ewa, Renata Tomaszewska, and Aleksandra Pawlicka. "Conflicting 'Mother-Scientist' Roles. An Innovative Application of Basket Analysis in Social Research." *PLoS ONE*, no. 17 (10) (2022). <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0276201>.
- Kurki, Visa A.J., and Tomasz Pietrzykowski, eds. *Legal Personhood: Animals, Artificial Intelligence and the Unborn*. Cham: Springer, 2017.
- Lipowicz, Markus. "Człowieczeństwo jako (nie)zbędna kategoria refleksji pedagogicznej? O ponowoczesnym kryzysie teorii wychowania w obliczu wyzwania transi posthumanizmu." *Studia z Teorii Wychowania* 8, no. 2 (19) (2017): 35–57.

- Omohundro, Steve. "Punkt zwrotny w dziedzinie sztucznej inteligencji." In *A ty co sądzisz o myślących maszynach? Wizje przyszłości wybitnych umysłów ery sztucznej inteligencji*. Edited by John Brockman. Translated by K. Kubala. Warszawa: Wydawnictwo Naukowe PWN, 2020.
- Pawlicka, Aleksandra, Marek Pawlicki, and Michał Choraś. "Zawód roku 2020. Jak go zdobyć?" In *Praca i kariera w warunkach nowego millennium. Implikacje dla edukacji*. Edited by Renata Tomaszewska. Bydgoszcz: Wydawnictwo UKW, 2022.
- Pawlicka, Aleksandra, et al. "Innovative Machine Learning Approach and Evaluation Campaign for Predicting the Subjective Feeling of Work-Life Balance Among Employees." *PLoS ONE*, no. 15 (5) (2020). <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0232771>.
- Pawlicka, Aleksandra, et al. "Has the Pandemic Made Us More Digitally Literate?" *Journal of Ambient Intelligence and Humanized Computing* (2022), <https://link.springer.com/article/10.1007/s12652-022-04371-1>.
- Przegalińska, Aleksandra K. *Istoty wirtualne: Jak fenomenologia zmieniała sztuczną inteligencję*. Kraków: Wydawnictwo Universitas, 2016.
- Roco, Mihail C., William Sims Bainbridge, National Science Foundation/Department of Commerce, eds. *Comerging Technologies for Improved Human Performance: Nanotechnology, Biotechnology, Information Technology and Cognitive Science*. New York: Kluwer Academic Publishers, 2003.
- Sejnowski, Terrence J. *Deep learning: Głęboka rewolucja: Kiedy sztuczna inteligencja spotka się z ludzką?* Translated by Piotr Cypryański. Warszawa: Wydawnictwo Poltext, 2019.
- Słownik, s.v. "Sztuczna inteligencja." <https://www.sztucznainteligenca.org.pl/definicja/sztuczna-inteligencja>.
- Walsh, Toby. *To żyje! Sztuczna inteligencja od logicznego fortepianu po zabójcze roboty*. Translated by Witold Sikorski. Warszawa: Wydawnictwo Naukowe PWN, 2018.
- Wodecki, Andrzej. *Sztuczna inteligencja w kreowaniu wartości organizacji*. Kraków–Legionowo: Wydawnictwo Edu-Libri, 2018.
- Wolfram, Stephen. *Sztuczna inteligencja i przyszłość cywilizacji*. In *Człowiek na rozdrożu. Sztuczna inteligencja – 25 punktów widzenia*. Edited by J. Brockman. Translated by Marcin Machnik. Gliwice: Wydawnictwo Helion, 2020.
- Zacher, Lech W. [Polska Akademia Nauk. Komitet Prognoz "Polska 2000 Plus"]. *Gry o przyszłe światy*. Warszawa: Warszawska Drukarnia Naukowa PAN, 2006.
- Zawojski, Piotr. *Technokultura i jej manifestacje artystyczne: Medialny świat hybryd i hybrydyzacji*. Katowice: Wydawnictwo Uniwersytetu Śląskiego, 2016.
- Zybertowicz, Andrzej, et al. *Samobójstwo Oświecenia? Jak neuronauka i nowe technologie pustoszą ludzki świat*. Kraków: Wydawnictwo Kasper, 2015.

ABSTRAKT / ABSTRACT

Renata TOMASZEWSKA – Raj – piekło – triumf. Sztuczna inteligencja w scenariuszach przyszłości

DOI 10.12887/36-2023-3-143-14

Celem podjętych rozważań jest ukazanie wybranych aspektów problemu badawczego sformułowanego w postaci pytania: Jak oswoić sztuczną inteligencję? W oparciu o metody krytycznej analizy literatury specjalistycznej i weryfikacji dokumentów formalno-prawnych wyjaśniono rozumienie pojęć „sztuczna inteligencja” i „superinteligencja”. W ich kontekście przytoczono trzy scenariusze przyszłości – rajski, piekielny i triumfalny, u podstaw których leży pytanie: jak oswoić sztuczną inteligencję? Podjęte zagadnienia badawcze wpisują się w dyskurs naukowy na temat sztucznej inteligencji i będą kontynuowane.

Słowa kluczowe: człowiek, sztuczna inteligencja, superinteligencja, przyszłość

Kontakt: Katedra Pedagogiki Pracy i Andragogiki, Wydział Pedagogiki, Uniwersytet Kazimierza Wielkiego, ul. Chodkiewicza 30, 85-064 Bydgoszcz

E-mail: renatatl@ukw.edu.pl

Tel. 52 3419256

<https://www.ukw.edu.pl/intranet/mojeDane.php>

https://www.ukw.edu.pl/pracownicy/strona/tomaszewska_lipiec//

Renata TOMASZEWSKA, Paradise—Hell—Triumph: Artificial Intelligence in Scenarios for the Future

DOI 10.12887/36-2023-3-143-14

The aim of the analysis is to show selected aspects of the category of artificial intelligence perceived as a research problem. Based on the methods of critical analysis of specialist literature and analysis of formal and legal documents, the understanding of the terms “artificial intelligence” and “superintelligence” has been clarified. In this context, the article presents three scenarios in which the future is perceived as, respectively, heaven, hell, and a triumph. Each of the scenarios is based on an answer to the question how to tame artificial intelligence. The issues considered in the paper are part of scholarly discourse on artificial intelligence and will be continued.

Translated by *Aleksandra Pawlicka*

Keywords: human, artificial intelligence, superintelligence, future

Contact: Department of Labour Pedagogy and Andragogy, Faculty of Pedagogy, Kazimierz Wielki University, ul. Chodkiewicza 30, 85-064 Bydgoszcz, Poland

E-mail: renatatl@ukw.edu.pl

Phone: +48 52 3419256

<https://www.ukw.edu.pl/intranet/mojeDane.php>

https://www.ukw.edu.pl/pracownicy/strona/tomaszewska_lipiec/

MYŚLAĆ OJCZYZNA...

Marek BERNACKI

„CZAS PRZYSZŁY JEDNĄ OGARNIE NAS NAZWĄ
NIEDOSKONAŁYCH I JUŻ ZAPOMNIANYCH”
O nieznanych juvenilnych wierszach Czesława Miłosza

Dlaczego trzy omawiane w tym artykule wiersze ujrzały światło dzienne tak późno? Jak doszło do tego, że utwory te „utknęły” na lat kilkadziesiąt w archiwum Teodora Bujnickiego? Czy była to świadoma decyzja starszego kolegi, aby tych młodzieńczych wierszy Miłosza nie publikować ze względu na ich niedoskonałość? A może było to redaktorskie przeoczenie lub zapomnienie? Prawdziwej przyczyny tego „zaniechania” zapewne już nigdy nie poznamy. Pozostaje cieszyć się z faktu, że trzy juvenilne wiersze wileńskiego poety, który marzył o sławie i tę sławę poetycką zdobył po wielu latach ciężkiej pracy, odnalazły się ku radości badaczy i miłośników jego twórczości.

Choć od śmierci Czesława Miłosza minęło już niemalże dwadzieścia lat, a twórczość poetycka noblisty, obejmująca około dziewięciuset wierszy i poematów powstałych w latach 1928-2003, zgromadzona w pięciu tomach *Wierszy* wydanych w edycji *Dzieł zebranych* oraz w monumentalnym tomie *Wiersze wszystkie*¹, została gruntownie opisana i omówiona przez niezliczone zastępy „badaczy owadzych nogów”², to jednak od czasu do czasu pojawiają się zaskakujące informacje o odkryciu nieznanych wcześniej i niepublikowanych utworów autora *Traktatu poetyckiego*. Tak było w przypadku wiersza *Piękna jesteś ojczyzno moja, Europo...*, odkrytego w Beinecke Rare Book and Manuscript Library (Yale University) przez Pawła Bema i opublikowanego w roku 2017 na łamach „Zeszytów Literackich”, a następnie przedrukowanego w „Gazecie Wyborczej”³, czy też paru innych wierszy Miłosza włączonych do

¹ Zob. C. Miłosz, *Wiersze*, t. 1-5, Znak, Kraków 2001-2009; t e n ż e, *Wiersze wszystkie*, Znak, Kraków 2011. Kolejne, uzupełniane na bieżąco wydania *Wierszy wszystkich* miały miejsce w latach: 2015, 2018 i 2021. W nocy *Od wydawcy* do wydania trzeciego, uzupełnionego z roku 2018 znajduje się następująca informacja: „Pierwsze wydanie *Wierszy wszystkich* ukazało się w 2011 roku, a w roku 2015 wydanie drugie, uzupełnione. Dodano wtedy kilka utworów opublikowanych już po przygotowaniu pierwszej edycji, zastosowano trochę inny układ wierszy rozproszonych oraz poprawiono zauważone błędy” (tamże, s. 1381).

² Wedle ironicznego określenia Czesława Miłosza, którym lubił się posługiwać, rozmawiając o pracach krytycznoliterackich poświęconych jego (i nie tylko jego) twórczości.

³ Por. C. Miłosz, *Piękna jesteś ojczyzno moja, Europo...*, [podał do druku P. Bem], „Magazyn Świąteczny Gazety Wyborczej”, 1-2 kwietnia 2017 roku, s. 33. Wiersz ten został włączony do trzeciego, uzupełnionego wydania *Wierszy wszystkich* (zob. t e n ż e, *Wiersze wszystkie*, Znak, Kraków 2018).

rozszerzonego, drugiego wydania wspomnianych wcześniej *Wierszy wszystkich*⁴. Niewątpliwie sporą niespodzianką lat ostatnich stało się wydobywanie na światło dzienne kilku rękopisów niepublikowanych, wczesnych wierszy i maszynopisów innych utworów oraz artykułów Miłosza, które były w posiadaniu Teodora („Dorka”) Bujnickiego. Przypomnijmy, bliskiego przyjaciela poety z czasów wileńskiej młodości i jednego z założycieli grupy poetyckiej Żagary⁵. Wspomniane rękopisy zostały przekazane 30 czerwca 2021 roku (dokładnie w sto dziesiątą rocznicę urodzin Miłosza) na ręce Krzysztofa Czyżewskiego, dyrektora Ośrodka „Pogranicze – sztuk, kultur, narodów” przez profesora Tadeusza Bujnickiego, wilnianina (urodzonego 14 lutego 1933 roku), syna Teodora, wybitnego literaturoznawcę i badacza przedwojennego życia literackiego w mieście nad Wilią i Wilejką.

Oddajmy głos profesorowi Tadeuszowi Bujnickiemu: „W papierach, które pozostały po moim ojcu Teodorze Bujnickim znalazło się kilka rękopisów pisanych ręką Czesława Miłosza oraz maszynopis utworu, który w odmiennej wersji został opublikowany w powojennym tomie poety *Ocalenie*. W posiadaniu Teodora Bujnickiego – poety grupy Żagary – znalazły się one nieprzypadkowo. Obu poetów łączyła wówczas przyjaźń oraz wspólny start w Żagarach (który to tytuł wymyślili wspólnie). Z tym, że ojciec był także redaktorem pisma, a wcześniej pełnił funkcję kierownika literackiego w «Alma Mater Vilnienisis», rocznika Uniwersytetu Stefana Batorego. Stąd też w jego rękach znalazły się: artykuł *Wyznanie jednego z młodych* oraz pięć wierszy. [...] Wymienione na początku wiersze i artykuł pochodzą z lat 1930-1931 i są juvenaliami Miłosza. Pisany ołówkiem wiersz *Podróż* to jeden z debiutanckich wierszy (obok *Kompozycji*) zamieszczony w 9 zeszytce «Alma Mater Vilnensis» z 1930 roku; natomiast wiersz *Tabu* drukują «Żagary» (1931 nr 3). T r z y p o z o s t a ł e w i e r s z e: *R a c h u n e k z n o w u*, *Z d o ł u* i *Z p e ł n ą ś w i a d o m o ś c i ą* ... (i n c i p.) n i e b y ł y p u b l i k o w a n e. [...] Na uwagę zasługuje sześciostronicowy artykuł *Wyznania jednego z młodych*, który warto czytać w kontekście znanego wiersza *Wam* ze słynnym zwrotem «z przekłętą jesteśmy bowiem pokolenia». Jest to próba określenia światopoglądu «młodych» o podobnym do dwudziestoletniego autora doświadczeniu życiowym»⁶. Przedmiotem naszych rozważań będzie analityczno-interpretacyjny ogląd trzech niepublikowanych dotąd wierszy Miłosza napisanych przez niego w latach 1930-1931: *Rachunek znowu*, „*Z pełną świadomością...*” oraz *Z dołu*.

⁴ T e n ż e, *Wiersze wszystkie*, Znak, Kraków 2015.

⁵ Grupa poetycka „Żagary” powstała w Wilnie w 1931 roku. Należeli do niej: Teodor Bujnicki, Henryk Dębiński, Stefan Jędrzychowski, Czesław Miłosz i Jerzy Zagórski. Poezja żagarystów sytuowana była w nurcie tzw. Drugiej Awangardy.

⁶ Por. T. B u j n i c k i, *Rękopisy Miłosza*, „Kroniki według Muzeum im. Marii Konopnickiej w Suwałkach” 2021, nr 4(29), s. 20 [wyróżnienie – M.B.].

Celowo zmieniam kolejność zaproponowaną przez Tadeusza Bujnickiego, co uwarunkowane jest tematyką wierszy. Dwa pierwsze teksty, podobne do siebie pod względem zarówno treści, jak i formy, uznać można za utwory o charakterze programowo-autotematycznym. Natomiast wiersz trzeci, zatytułowany *Z dołu*, to dość sprawnie napisany, intrygujący pod wieloma względami, poetycki obrazek rodzajowy. Zgodnie z sugestią Bujnickiego warto popatrzeć na te trzy wczesne utwory Miłosza, jak na „dokument kształtowania się warsztatu i artystycznej wrażliwości” przyszłego noblisty⁷.

Zanim przejdę do analizy wymienionych tekstów, opublikowanych po raz pierwszy jesienią 2021 roku na łamach suwalskich „Kronik”, redagowanych przez Zbigniewa Fałtynowicza⁸, przypomnę pokrótce biograficzne i historyczne konteksty, w jakich te młodzieńcze wiersze powstawały. W roku 1930, kiedy na łamach uniwersyteckiego rocznika „Alma Mater Vilnensis” ukazały się dwa debiutanckie utwory Miłosza – *Kompozycja* i *Podróż*, ich autor miał zaledwie dziewiętnaście lat. Mieszkał w Wilnie, do którego przyjechał w 1921 roku z nadniewieskich Szetejni z rodzicami i młodszym bratem Andrzejem. Po ukończeniu ośmioletniego Gimnazjum im. Zygmunta Augusta, w którym po raz pierwszy – jak to ujął Andrzej Zawada – głęboko doświadczył sporu między humanizmem a służbą jednej ideologii⁹, osiemnastoletni Miłosz zapisał się na Uniwersytet Stefana Batorego. Zaledwie przez kilka tygodni studiował polonistykę, by ostatecznie przenieść się na Wydział Prawa, który ukończył w 1934 roku, otrzymując dyplom magisterski.

Mieszkając w Wilnie, na przełomie lat dwudziestych i trzydziestych dwudziestego wieku, początkujący poeta doświadczył klimatu tego niepowtarzalnego miasta, zwanego niekiedy Jerozolimą Północy. Całym sobą chłoniął niepowtarzalną atmosferę stolicy byłego Wielkiego Księstwa Litewskiego z jego nadal obecną i wyczuwalną wieloetnicznością oraz różnorodnością językową, kulturową czy wyznaniową. Od wczesnej młodości Miłosz zdecydowanie miał lewicowe poglądy. Irytowały go antysemickie burdy organizowane przez Młodzież Wszechpolską, nie tolerował ideologii lansowanej przez obóz władzy i endecką propagandę. Już w gimnazjum postanowił wstąpić do utajnionej organizacji uczniowskiej PET (Związek Młodzieży Polskiej „Przyszłość”), będącej czymś w rodzaju młodzieżowej przybudówki popularnych w Wilnie

⁷ Por. tamże.

⁸ Por. Z. Fałtynowicz, *Dar*, „Kroniki według Muzeum im. Marii Konopnickiej w Suwałkach” 2021, nr 4(29), s. 20-25.

⁹ Por. A. Zawada, *Miłosz*, Wydawnictwo Dolnośląskie, Wrocław 1996, s. 25. Chodzi o spór światopoglądowy prowadzony między filologiem klasycznym Settembrinim a jezuitą Napthą, którzy walczyli o duszę Hansa Castorpa, głównego bohatera powieści Tomasza Manna *Czarodziejska góra* (por. np. T. Mann, *Czarodziejska góra*, t. 2, tłum. W. Tatarkiewicz, Świat Książki, Warszawa 1998, s. 43-62).

stowarzyszeń masońskich, lansujących wartości liberalne i kosmopolityczne¹⁰. W latach studiów młody Miłosz, mimo że oficjalnie studiował prawo, chętnie uczęszczał na wykłady historiozoficzne profesora Mariana Zdziechowskiego, słuchał także wystąpień znanych wileńskich polonistów – profesora Manfreda Kridla¹¹ i profesora Stanisława Pigoń¹². Niezwykle aktywnie udzielał się w Sekcji Twórczości Oryginalnej (STO), działającej w ramach Koła Naukowego Polonistów (opiekunem koła był profesor Pigoń). To właśnie wtedy, czyli na przełomie 1929 i 1930 roku, powstały jego pierwsze utwory poetyckie, o których donosił Jarosławowi Iwaszkiewiczowi w liście, prosząc znanego skamandrytę o opinię: „Z wierszy, które posyłam, *Kompozycja* i *Podróż* będą drukowane wkrótce w «Alma Mater Vilnensis». Ale to wszystko głupstwo i póki Pan nie wyda swojej oceny, nic nie wiem o swoich tworach”¹³.

Renata Gorczyńska zauważa: „Miłosz ruszył długą drogą poezji jeszcze w gimnazjum, co poniekąd typowe, uprawiając wprawki mimetyczne w stylu klasycznej poezji francuskiej. Zdarzało mu się też naśladować młodego Mickiewicza, którego duch po dziś dzień unosi się nad miastem. Jak wspomina, «w okresie prehistorycznym», mieszkając w klasie maturalnej na stacji u pani Kłęckiej na przełomie lat 1928/29 w otoczeniu starszych od siebie młodzieńców, już studentów, pisywał wiersze w zeszytach, a jak zabrakło w nich miejsca, to na skrawkach papieru. Był to, jak określił, rok jego wielkich oczekiwań. Coś w nim dojrzewało i domagało się ujścia. Zaczął się ów proces jesienią, ulubioną porą Miłosza, w jego oczach szczególnie piękną porą roku w Wilnie. Gdy chodził drewnianymi chodnikami wąskich uliczek, spodziewał się jakiegoś potężnego przełomu w swoim życiu, jak mi opowiadał”¹⁴. Na lata 1928-1930 przypada początek kształtowania się światopoglądu i światoodczucia poetyckiego przyszłego autora *Trzech zim*. Ożywione dyskusje w kręgu młodych literatów, takich jak Teodor Bujnicki i Kazimierz Hałaburda, czy poznani w Akademickim Klubie Włóczęgów Stefan Jędrzychowski i Stefan Zagórski, a nieco później także Jerzy Zagórski i Henryk Dębiński – zaowocują powstaniem poetyckiej grupy Żagary, która odegrała bardzo ważną rolę w roz-

¹⁰ Jak podaje Andrzej Franaszek, cytując opracowania źródłowe: „Wileński PET kultywował liberalizm, tolerancję i z niechęcią odnosił się do nacjonalizmu” (A. Franaszek, *Miłosz. Biografia*, Znak, Kraków 2011, s. 94). Co ciekawe, do wspomnianej organizacji należeli obok Miłosza także Teodor Bujnicki, Antoni Gołubiew, Czesław Zgorzelski czy Stanisław Stomma.

¹¹ Prof. Manfred Kridl (1882-1957), historyk literatury, autor syntez historycznoliterackich poświęconych literaturze polskiej i europejskiej XIX i XX wieku.

¹² Prof. Stanisław Pigoń (1885-1968), historyk literatury polskiej, wybitny znawca twórczości Adama Mickiewicza, edytor i pedagog.

¹³ C. Miłosz, J. Iwaszkiewicz, *Portret podwójny. Wykonany z listów, wierszy, zapisków intymnych, wywiadów i publikacji*, wybór B. Toruńczyk, oprac. R. Papieski, Fundacja Zeszytów Literackich, Warszawa 2011, s. 7.

¹⁴ R. Gorczyńska, *Miłosz wileński. Część II*, „Kwartalnik Artystyczny” 2021, nr 4, s. 75.

woju polskiej poezji okresu dwudziestolecia międzywojennego. Z tego tygla młodzieńczego buntu i „nieba w płomieniach”, literackich pasji i poszukiwań, wczesnej lektury pism Stanisława Brzozowskiego¹⁵, społeczno-politycznych dyskusji toczących się w okresie rewolucyjnego wrzenia i rozpoczynającego się w Europie wielkiego kryzysu ekonomicznego¹⁶, pierwszej, szalonej zagranicznej wyprawy przez Czechosłowację i Niemcy do Paryża, a także bujnego okresu dojrzewania naznaczonego licznymi flirtami i romansami („Lata odmierzane różnymi gatunkami kobiet przemijały”¹⁷) – zrodził się pierwszy tom poetycki Miłosza *Poemat o czasie zastygłym*¹⁸. Według Górczyńskiej: „*Poemat...* poematem nie był. Był natomiast wyborem niespełna piętnastu wierszy, w których znalazło się dużo więcej polityki niż metafizyki. Dwudziestodwuletni autor nie szczędził gorzkich słów pod adresem rządców II RP za niesprawiedliwość społeczną, szowinizm, tromtadrację, obojętność na dolę prostego człowieka i militaryzm na pokaz. Jego gniew był w dużej mierze słuszny, ale nie służył zbyt dobrze spadkobiercy lutni po Bekwarku. Niemniej jednak i wiersze własne, i te zebrane w antologii, też po latach przez niego samego skrytykowane, świadczyły o zasadniczej niezgodzie na porządek świata”¹⁹. W tym też czasie powstały trzy wiersze, nigdzie dotąd niepublikowane, które przeleżały w prywatnym archiwum Teodora Bujnickiego ponad dziewięćdziesiąt lat, by w czerwcu 2021 roku ujrzeć w końcu światło dzienne. Przyjrzymy się najpierw dwóm pierwszym juvenilnym poetyckim tekstom Miłosza – zaczynającemu się od incipitu utworowi „*Z pełną świadomością...*”, a następnie wierszowi zatytułowanemu *Rachunek znowu*:

Z pełną świadomością czego pragnąć warto
Tworzę te rzeczy piękne i znikome
Nie mogąc znaleźć innego spokoju
W trudnej epoce wielkiego przełomu.

Zarówno słowa, teorii, zjawiska
Nawpół pojmując, jednak prawidłowo
Łączyłem, wiedząc że bardziej od prawdy
Potrzebna wola kochająca nowość.
A to nie znaczy, żebym mocno wierzył

¹⁵ Jak podaje Marta Wyka, Miłosz w wieku dwudziestu lat zainteresował się pismami Brzozowskiego i pilnie studiował jego *Legendę Młodej Polski* (por. M. Wyka, *Miłosz i rówieśnicy. Domknięcie formacji*, Universitas, Kraków 2013, s. 56).

¹⁶ „Spotykają się najczęściej w kawiarni Rudnickiego przy ulicy Mickiewicza, dziś bulwarze Gedymina. Ten i inne lokale stają się areną ich zażartych dysput o poezji, ale i o polityce” (zob. Górczyńska, dz. cyt., s. 76).

¹⁷ Zob. C. Miłosz, *Opowieść*, w: tegoż, *Wiersze*, t. 1, Znak, Kraków 2001, s. 9.

¹⁸ Zob. tenże, *Poemat o czasie zastygłym*, Koło Polonistów Uniwersytetu Wileńskiego, Wilno 1933.

¹⁹ Górczyńska, dz. cyt., s. 81.

W wieczne znaczenie myśli zapisanych.
Czas przyszły jedną ogarnie nas nazwą:
Niedoskonałych i już zapomnianych.

Wobec tak wielkiej przemienności dziejów
Równe są czyny ludzkie i zwierzęce
I nikt nie pyta, czy na wąskich grobach,
Trawa wyrośnie, czy blaszane wieńce.

Mimo dążenia, trudno jest wyprzedzić
Epokę, albo stanąć u jej granic.
Więc obowiązek sprawiał że pisałem
O rzeczach które zwycięży niepamięć²⁰.

I wiersz drugi:

słowa ciężkie i męskie napróżno pisałem
barwami i rytmem napróżno kołysałem
wiem że jestem skazany na płomień bezmierny
jak brąz tak ukochany i jak czerń tak wierny

przed ołtarzem obrazów muszę leżeć krzyżem
w zapachu ciężkim świata w locie światła chyżym
i widzieć jak ekrany palą się i pienia
i widzieć jak rubinem zachodzą przestrzenie

wszystkie prawdy drapieżne tak długo noszone
wszystkie szczerze modlitwy przed śmierci obliczem
bóstwu oddam mojemu i już chyba ciszy
nazbyt słodkiej nie przetną zgrzyty ciężkich powiek

podły jestem jezu podły jestem jezu, jezu, jezu

jak krzyk ptaków czerwonych wykwitają nowe
kolory rdzawej bieli jak śpiewy cięciwy
ja płasam jak tancerka i podniósłszy głowę
w scenie serc poszarpanych wstrząsam złoto grzywy²¹

W utworze „*Z pełną świadomością...*” na plan pierwszy wysuwa się kwestia świadomości poetyckiej podmiotu lirycznego odczuwającego wewnętrzną

²⁰ C. Miłoś, „*Z pełną świadomością...*” [rękopis], cyt. za: „Kroniki według Muzeum im. Marii Konopnickiej w Suwałkach” 2021, nr 4(29), s. 21. We wszystkich wierszach spisywanych z rękopisów Miłosa wydrukowanych w formie fotokopii na łamach suwalskich „Kronik” zachowuję oryginalną interpunkcję i ortografię zgodną z pisownią autora.

²¹ C. Miłoś, *Rachunek znowu* [rękopis], cyt. za: „Kroniki według Muzeum im. Marii Konopnickiej w Suwałkach” 2021, nr 4(29), s. 23.

powinność utrwalania w słowie „rzeczy pięknych i znikomych” w niepewnych i nietrwałych czasach „wielkiego przełomu” i „przemienności dziejów”. Podmiot wiersza podkreśla jednocześnie wagę historyczno- i teoretycznoliterackiego warsztatu poetyckiego, co wyrażone zostaje przez metonimiczne określenia: „słowa, teorie, zjawiska”, które oddają zapewne klimat toczących się w wileńskich kawiarniach dyskusji o życiu literackim międzywojennej Polski i Europy. Jednocześnie młody poeta wyraźnie zaznacza, że bardziej niż kurczowe przywiązanie do literackiej tradycji pociąga go tropienie nowości, co oznaczało w praktyce uporczywe poszukiwanie własnego stylu i wykuvanie poetyckiego idiolektu. Kolejną rzeczą, na którą warto zwrócić uwagę, analizując wiersz „*Z pełną świadomością...*”, jest – nomen omen – dotkliwa świadomość, czy raczej przecucie nadciągającej klęski, fatalistyczne przekonanie podmiotu lirycznego o daremności wszelkich podejmowanych przez niego i jego przyjaciół starań, które zostaną zniweczone przez nadchodzące wydarzenia: „Czas przyszły jedną ogarnie nas nazwą / Niedoskonałych i już zapomnianych”. Znamienne jest, że pisząc ten dwuwiersz, Miłosz niejako prefiguruje późniejsze fatalistyczne frazy o charakterze pokoleniowym, wypowiedziane kilkanaście lat później przez młodszych od niego o dekadę poetów pokolenia Kolumbów (rocznik 1920) – Krzysztofa Kamila Baczyńskiego i Tadeusza Borowskiego: „Zostanie po nas złom żelazny / i głuchy, drwiący śmiech pokoleń”²². Fatalistyczny ton widoczny jest także w drugim z przytoczonych wierszy pod tytułem *Rachunek znowu*. Już sam tytuł jest znamieny – in medias res wprowadza czytelnika w klimat młodzieńczego rozrachunku prowadzonego przez młodego poetę z samym sobą. Twórczość poetycka jawi się w tym oryginalnym utworze jako ofiara składana przez autora na ołtarzu obrazów, barw i rytmów powiązanych w jednolitą strukturę słowną. Co ciekawe, w odróżnieniu od pierwszego z analizowanych wierszy mamy tu do czynienia ze sporym nasyceniem obrazowania wizyjno-katastroficznego, które poeta rozwinie na szeroką skalę w bardziej dojrzałych utworach z tomu *Trzy zimy*²³. I tak klimat nadciągającego, przeczuwanego przez „ja” liryczne jakiegoś wielkiego zagrożenia oddany został nade wszystko przez odniesienie do obrazu „płomienia bezmiernego” oraz sekwencję hermetycznych wyrażań metaforycznych: „zapach ciężki świata”, „lot światła chyży”, palące się i pieńiące tajemnicze „ekrany” czy „rubinem zachodzące przestrzenie”.

²² Końcowy dwuwiersz *Pieśni* Tadeusza Borowskiego, będącej fragmentem cyklu poetyckiego tegoż autora *Gdziekolwiek ziemia...* – poetyckiego debiutu opublikowanego w 1942 roku podczas okupacji hitlerowskiej (por. „*Na wrywyki*”. 100 cytatów z polskiej poezji i dramatu, które powinien znać także cudzoziemiec, red. R. Cudak, W. Hajduk-Gawron, A. Madeja, Wydawnictwo Uniwersytetu Śląskiego, Katowice 2018, s. 367n.).

²³ Zob. C. M i ł o s z, *Trzy zimy*, Związek Zawodowy Literatów Polskich, Wilno–Warszawa 1936.

Uwagę zwraca ponadto jeden, wydzielony przez poetę wers, umieszczony między trzecią i czwartą (finalną) zwrotką: „podły jestem jezu podły jestem jezu jezu jezu”. Ta ekspresjonistyczna i manieryczna zarazem inkantacja, świadcząca o znacznym rozedrganiu emocjonalnym autora wiersza (przypomnijmy: osiemnastoletniego, może dziewiętnastoletniego w chwili pisania tego tekstu), przywodzi na myśl fragment debiutanckiej *Kompozycji*, wiersza opublikowanego przez Miłosza w 1930 roku na łamach „Alma Mater Vilnensis”:

dzwony dzwony wybuchły
pons christi pons christi
pons pons christi

rozdarły się kurtyny z trzaskiem
patrzcie – jesteście w teatrze²⁴

Sytuacja liryczna *Kompozycji* osadzona została najprawdopodobniej we wnętrzu wileńskiego barokowego kościoła Świętych Apostołów Piotra i Pawła na Antokolu. Jest to utwór o młodzieńczym kryzysie wiary, zaś zacytowany fragment przełamuje niejako przestrzeń wiersza, oddzielając strefę sacrum od strefy profanum (ta pierwsza jest usytuowana wewnątrz świątyni, ta druga – na deskach teatru, który symbolizuje życie w jego różnych, także tych grzesznych, nieuporządkowanych przejawach). W badanym przez nas nieznanym wcześniej wierszu Miłosza mamy do czynienia z podobnym zabiegiem, z tym, że zamiast epatować czytelnika problemem „nieba w płomieniach”, młody poeta, porównując swe zachowanie do „płasów tancerki na scenie”, w ironiczny sposób przedstawia kryzys twórczy, posługując się sekwencją mniej lub bardziej udanych ekwiwalentów poetyckich:

jak krzyk ptaków czerwonych wykwitają nowe
kolory rdzawej bieli jak śpiewy cięciwy
ja płasam jak tancerka i podniósłszy głowę
w scenie serc poszarpanych wstrząsam złoto grzywy²⁵

Przyjrzyjmy się teraz trzeciemu z zapowiedzianych utworów poetyckich, zatytułowanemu *Z dołu*:

Asfalt lekko chropawy. dwa buty to skały.
szybki jak obrót szprych przepływa prąd cienia

²⁴ T e n ż e, *Kompozycja*, w: tenże, *Wiersze*, t. 1, s. 31. Wiersz ten szerzej analizuję w książce *Tropienie Miłosza* (por. M. B e r n a c k i, *Tropienie Miłosza. Hermeneutyczna „bio-grafia” Poety*, Universitas, Kraków 2019, rozdz. „Młodzieńczy kryzys wiary („Kompozycja”)”, s. 9-15).

²⁵ M i ł o s z, *Rachunek znowu*, dz. cyt., s. 23.

moneta światła czasem upadnie tak blisko
że z ust poszukiwacza złota wyrwałby się krzyk zachwycenia.

wyżej są zmięte słupy spodni. ich kieszenie
naciągnięte pod ciężarem kwadratowych pięści.
w jednej jest brudna chustka pełna krwi i plwocin
w drugiej 30 groszy niespokojnie chrzęści.

przybawajcie o kule. Kamizelka kryje
brązowe punkty piersi a pod jednym z nich serce słabo bije tik-tak
aorta jak armata wyrzuca pocisk gorącego płynu
w różowych płucach leżą niteczki karminu
piers zarośnięta broni motoru od ciosu.

krawat nie jest tkaniną. To sieć zaciągnięta
w węzeł. gdy głowa w matni miotać się zaczęła
oczka zwarły się. oto jak ryba bez głosu
głowa cierpi łapczywie ssie dym z papierosów.

magnes oczu przyciąga bliżej prostokąty
zielone autobusów, romby czerwone taksówek
zygzak reklamy szybko idzie po średnicy
potem zamknie półkole i spływa w dół strzałą.

niebo czarne nad tobą jak róża dojrzało
ale zerwać nie można bo jest za wysoko

choćby nawet lady dała 3 dolary²⁶.

Wiersz ten zdecydowanie różni się od dwóch utworów przytoczonych i omówionych wcześniej. W przeciwieństwie do nich należy do liryki pośredniej. Można powiedzieć, że jest poetyckim portretem anonimowego człowieka, który znalazł się w polu obserwacji autora wiersza. Oryginalny jest przy tym niejako geometryczny sposób patrzenia na prezentowaną osobę²⁷ – bliżej nieokreślonego mężczyznę, zapewne bezrobotnego proletariusza. Obserwator, opisujący tę postać, przyjmuje tak zwaną żabią perspektywę, posługując się metodą oddolnego spojrzenia, o czym informuje tytuł utworu. Patrzenie „z dołu” (a nie „z góry”!) ma przy tym ukrytą wymowę moralną, sygnalizuje niejako próbę utożsamienia się autora z bohaterem wiersza, przyjęcie wobec niego postawy empatycznej. Taki sposób spoglądania na kogoś czy na coś, znany prekursorom malarstwa nowoczesnego Gustawowi Klimtowi

²⁶ C. Miłoś z, *Z dołu* [rękopis], cyt. za: „Kroniki według Muzeum im. Marii Konopnickiej w Suwałkach” 2021, nr 4(29), s. 22.

²⁷ Szerzej na temat geometryzacji przestrzeni w poezji żagarystów zob. M. Zalecki, *Przygoda drugiej awangardy*, Zakład Narodowy im. Ossolińskich, Wrocław 2000.

czy Henriemu Toulouse-Lautrecowi, sprawia, że obserwowany podmiot (lub przedmiot) staje się niejako większy, a tym samym godny uwagi. Kim zatem jest ów obserwowany przez poetę nieznajomy? Trudno jednoznacznie określić. Z treści wiersza wywnioskować można, że najprawdopodobniej jest robotnikiem, który został wyrzucony na bruk i nie wie, co ma ze sobą zrobić. Być może jest przedstawicielem lumpenproletariatu, z tajemniczą, kto wie, czy nie kryminalną, przeszłością. Sporo przemawia za tym, że to człowiek z marginesu społecznego, opuszczony i schorowany (najpewniej na gruźlicę – czyżby nabawił się jej w więzieniu?). Z opisu wynika, że mężczyzna jest pozbawiony nadziei, przygnębiony i osaczony, o czym świadczyć może ekspresjonistyczny opis krawata porównanego przez poetę do pętli oplatającej szyję protagonisty wiersza („sieć zaciągnięta w węzeł”). Człowiek ten najwyraźniej przymiera głodem, a jego codziennym „pożywieniem” są tanie papierosy, których palenie wyznacza mu, obok obserwowania wydarzeń rozgrywających się na ulicy, rytm dnia i nocy. Dopelnieniem takiego oddolnego, wertykalnego sposobu patrzenia na owego nieszczęśnika jest finalny, metaforyczny obraz „czarnego nieba, które dojrzewa różą”. Kwiat ten symbolizuje wartość nieodstępną dla bohatera wiersza skazanego na skromne finansowe datki przechodniów, czego wymownym przykładem są trzy dolary rzucone przez zamożną lady.

Wiersz *Z dołu* określić można jako utwór zaangażowany ideowo, wyraz społecznych, zdecydowanie lewicowych fascynacji młodego Miłosza. Jest to utwór wyrastający z klimatu pokoleniowych dyskusji światopoglądowych przyszłych żagarystów, krytycznie nastawionych do prawicowych rządów II Rzeczypospolitej, ale też wynik obserwacji początkującego literata, a zarazem młodego i gniewnego przedstawiciela wstępującej generacji, któremu nie było przecież obce doświadczenie ludzkiej nędzy widocznej gołym okiem w Wilnie. Przed wojną Wilno było jednym z najbiedniejszych polskich miast borykających się ze skutkami gospodarczego i politycznego kryzysu początku lat trzydziestych dwudziestego wieku²⁸.

Podsumujmy dotychczasowe rozważania. Analiza trzech niepublikowanych wcześniej juvenilnych wierszy Miłosza nie przynosi zaskakujących odkryć. Można powiedzieć, że zarówno swoją formą, jak i wymową ideową utwory te wpisują się w tonację młodzieńczej poezji autora *Poematu o czasie zastygłym*. Podobnie jak wiersze z tego tomu czy utwory rozproszone, drukowane na łamach przedwojennych periodyków, napisane na początku lat trzydziestych, pokazują Miłosza jako człowieka wrażliwego na społeczną krzywdę, ale zarazem też indywidualistę, poszukującego własnej linii rozwoju, oryginalnej poetyckiej dykcji. Występujące w dwóch pierwszych tekstach wąt-

²⁸ Szerzej na ten temat zob. C. Miłosz, *Wyprawa w dwudziestolecie*, Wydawnictwo Literackie, Kraków 2000.

ki fatalizmu dziejowego ciężącego nad pokoleniem, do którego należał poeta, z dużo większą intensywnością pojawiają się w innych wierszach z tego okresu. Na przykład w *Domu młodzieńców* („Czy przeminie ziemia? Czy przeminie olbrzymi nieboskłon? Nie. Przeminać trzeba nam”²⁹) czy opublikowanym na łamach „Żagarów” gniewnym manifestie wstępującej generacji pod tytułem *Wam* („A my już upadamy. I jeszcze od wschodu / Obracamy swe oczy. Ale przecie ziemia / Zaspokoić naszego nie potrafi głodu / Z przekłętą bowiem jesteśmy pokolenia”³⁰). Osiągają zaś kulminację w wierszu *Na śmierć młodego mężczyzny* z debiutanckiego tomu *Poemat o czasie zastygłym*, któremu patronuje rosyjski poeta skandalista Sergiusz Jesienin:

Mając dwadzieścia lat i leżąc w ciemności późnej
Leniwi od snu o gwiazdach i siana świeżego stogach
Wpadamy czasem w wir i nagle ogarnia nas trwoga
I szybko bić zaczynają tętnice splecione na mózgu.

Widzimy wtedy nasz koniec. O świecie śnieg poranny
Napływa pod senne okna, szczelnie czymś czarnym zakryte –
Gromnice palą się mdło. Ktoś z książki modlitwy czyta –
A my chrapiemy, rżemy, długo i ciężko konamy³¹.

Z kolei ton antykapitalistyczny, który pojawia się w wierszu *Z dołu*, będący wyrazem empatycznego utożsamienia się z losem anonimowego lumpen-proletariusza, przenika niemal wszystkie utwory włączone przez Miłosza do tomu *Poemat o czasie zastygłym*³². Zbiór ten, podobnie jak *Antologia poezji społecznej*, wydana we współpracy ze Zbigniewem Folejewskim, czy młodzieńcza publicystyka Miłosza z wczesnego okresu jego twórczości³³, nosi

²⁹ C. Miłosz, *Dom młodzieńców*, w: tenże, *Wiersze*, t. 1, s. 25.

³⁰ Tenże, *Wam*, w: tenże, *Wiersze*, t. 1, s. 37.

³¹ Tenże, *Na śmierć młodego mężczyzny*, w: tenże, *Wiersze*, t. 1, s. 26.

³² Zob. tenże, *Poemat o czasie zastygłym*, Koło Polonistów Słuchaczy Uniwersytetu Stefana Batorego, Wilno 1933.

³³ Zob. *Antologia poezji społecznej 1924-1933* [za Komitet Wydawniczy Czesław Miłosz, Zbigniew Folejewski], Koło Polonistów Słuchaczy Uniwersytetu Stefana Batorego, Wilno 1933. Najbardziej radykalnym przykładem publicystyki zaangażowanej światopoglądowo jest artykuł pod tytułem *Bulion z gwoździ* opublikowany w grudniu 1931 roku na łamach „Żagarów”, w którym Miłosz w swym radykalizmie zbliża się do teorii Żdanowa, teoretyka sowieckiego socrealizmu. Oto próbka młodzieńczego stylu przyszłego noblisty: „Sztuka jest narzędziem w walce społeczeństw o wygodniejsze formy bytu. Jej rola polega na organizowaniu psychiki. [...] Producent dóbr artystycznych musi mieć jasny i wyraźny cel. Tym celem powinno być urobienie takiego typu człowieka, jaki ze względów społecznych jest w najbliższej przyszłości potrzebny. A więc artysta kieruje hodowlą ludzi” (cyt. za: C. Miłosz, *Bulion z gwoździ*, w: tenże, *Przegląd młodego umysłu. Publicystyka i proza 1931-1939*, red. A. Stawiarska, A. Fiut, Znak, Kraków 2003, s. 32n.).

znamiona frenezji rewolucyjnej i odzwierciedla młodzieńcze, mocno lewicujące, choć jeszcze nie marksistowskie zainteresowania dwudziestoletniego poety formułowane *expressis verbis* między innymi pod wpływem przyjaźni ze starszym kolegą z domu studenckiego, Litwinem Pranasem Ancevicusem (Franciszkiem Ancewiczem). Jak pisze Gorczyńska, był „rewolucyjnym socjalistą, współlokatorem i partnerem w dyskusjach”³⁴.

W utworach pisanych przez młodego Miłosza przykładów rewolucyjnej tonacji, przypominających agitacyjne wiersze Władysława Broniewskiego czy Włodzimierza Majakowskiego, jest sporo. Największe ich nasycenie występuje w wierszu *Przeciwko nim*, napisanym – jak informował w motcie do tego utworu sam autor – „z powodu procesu Blachowskiego, zabójcy Gastona Koehler-Badin”³⁵. Wiersz ten, będący oskarżeniem systemu kapitalistycznego, odpowiadającego według poety za nędzę polskich robotników i upodlenie ich rodzin, a także przyczyniającego się do tworzenia ogromnych fortun przez nieuczciwych właścicieli-krwiopijców, zaczyna się niczym partyjna agitka, skrojona według wzorców lansowanych przez barda rewolucji leninowskiej, Majakowskiego:

To jest ziemia płaska i uboga
– dla podróżnych z Moskwy do Berlina
kolonia francuskich kapitałów
eksploatowane tereny.

Nam ten kraj – wieńcem na karabinach.
My ten kraj
pięścią
podniesiemy.

Tu jest nędza. Tu jest krzyk głodnych mas.
Tu jest rozpacz, co ściska gardło.
... W ciemnych dymach i w groźny czas
dnem i nocą
huczał
Żyrardów³⁶.

³⁴ Gorczyńska, dz. cyt., s. 76. Zob. też V. Daujotytė, M. Kvietauskas, *Litewskie konteksty Czesława Miłosza*, tłum. J. Tabor, Pogranicze, Sejny 2014 (rozdz. „Na powojennych kontynentach: wileńska emigracja i ‘Literatūros lankai’”, s. 306-343).

³⁵ Julian Blachowski, pracownik Zakładów Żyrardowskich, które w tamtym czasie należały do kapitału francuskiego, zastrzelił na ulicy naczelnego dyrektora tychże Zakładów. Jego proces toczył się w październiku 1932 roku i zakończył się skazaniem polskiego robotnika na pięć lat więzienia (por. Miłosz, *Przemyśle młodego umysłu*, s. 247).

³⁶ Tenże, *Przeciwko nim*, w: tenże, *Wiersze*, t. 1, s. 14.

Na koniec kilka uwag o formie trzech juvenilnych wierszy Miłosza. Łatwo zauważyć, że w każdym z analizowanych tekstów dominuje model tradycyjnej poetyki, w większym stopniu odwołujący się do spuścizny romantycznej i neoromantycznej aniżeli do postulatów awangardowych proveniencji peiperowskiej. Analizowane utwory pisane są na ogół nieregularnym wierszem sylabicznym (jedenastozgłoskowym, dwunastozgłoskowym, trzynastozgłoskowym czy czternastozgłoskowym), z zachowaniem średniówki (sic!) i tradycyjnych rymów (czasami dość nieudolnie konstruowanych, szukanych przez początkującego poetę niejako na siłę). Pojawiają się też niezbyt przekonujące próby wprowadzenia wiersza sylabotonicznego z miarą stałym rozkładem akcentów³⁷. W dwóch pierwszych tekstach dominuje styl hieratyczny, podniosły, wzmacniany przez elementy retoryczne, dzięki którym młody poeta, uprawiając model liryki bezpośredniej, pisanej w pierwszej osobie liczby pojedynczej, pragnie nadać głębszy sens poruszonym w utworach kwestiom artystycznym, społecznym i historiozoficznym. Ponadto w wierszu *Rachunek znowu* pojawiają się, co już wcześniej zostało dostrzeżone, elementy poetyki wizyjno-symbolicznej, operującej metaforycznym, peryfrastycznym obrazowaniem charakterystycznym dla katastroficzno-eschatologicznych wierszy z tomu *Trzy zimy* (1936), po wydaniu którego Miłosza okrzyknięto najważniejszym poetą wstępującego pokolenia. W model liryki awangardowej najbardziej wpisuje się wiersz *Z dołu*, przypominający formalistyczno-kubistyczne studium anonimowego antybohatera, „produktu” systemu kapitalistycznego, wyalienowanego lumpenproletariusza³⁸. W trzech analizowanych wierszach Miłosz posługuje się nietypową dla ówczesnego modelu liryki popularnej pisownią z małej litery, bez konsekwencji w stawianiu znaków przestankowych, z pojawiającymi się tu i ówdzie przerzutniami, inwersjami czy elipsami. Wszystkie te dostrzeżone i opisane w sposób ogólny elementy poetyki świadczą o mozolnym procesie poszukiwania i wypracowywania przez Miłosza własnej, oryginalnej formy wypowiedzi poetyckiej. Kończąc, pozostaje do wyjaśnienia jeszcze jedna, intrygująca niewątpliwie kwestia: dlaczego trzy omawiane w tym artykule wiersze ujrzały światło dzienne tak późno? Innymi słowy, jak doszło do tego, że utwory te „utknęły” na lat kilkadziesiąt w prywatnym archiwum Teodora

³⁷ Notabene, jak zauważył wiele lat później Stanisław Balbus, Miłosz sięgał po niemal wszystkie możliwe typy wiersza znane polskiej literaturze (zob. S. B a l b u s, „Pierwszy ruch jest śpiewanie” (*O wierszu Miłosza – rozpoznanie wstępne*), w: *Poznawanie Miłosza*, red. J. Kwiatkowski, Wydawnictwo Literackie, Kraków 1985).

³⁸ Wiersz ten nosi wyraźne piętno stylu Czechowiczowskiego z wczesnej fazy twórczości autora tomu *Dzień jak co dzień* (zob. J. C z e c h o w i c z, *Dzień jak co dzień*, f. hoesick, Warszawa 1930). Miłosz znał na pewno twórczość Józefa Czechowicza, jednego z filarów Drugiej Awangardy i członka lubelskiej grupy poetyckiej Reflektory. Od roku 1934 datuje się przyjaźń obu poetów, która trwała nieprzerwanie aż do tragicznej śmierci Czechowicza w pierwszych dniach września 1939 roku, kiedy zginął podczas niemieckiego nalotu na Lublin.

Bujnickiego? Czy była to świadoma decyzja starszego kolegi i mentora, aby te młodzieńcze wiersze Miłosza „odstawić” i nie publikować ze względu na ich niedoskonałość? A może wynikało to z jego redaktorskiego przeoczenia lub zapomnienia? Być może sprawę rozstrzygnęli między sobą obydwaj żagaryści, Teodor Bujnicki i Miłosz? Kto wie... Prawdziwej przyczyny tego „zaniechania” zapewne już nigdy nie poznamy. Pozostaje cieszyć się z faktu, że trzy dawno temu napisane wiersze, będące świadectwem chmurnej i dumnej młodości wileńskiego poety, który marzył o sławie i tę sławę poetycką zdobył po wielu latach ciężkiej pracy, odnalazły się ku radości badaczy i miłośników jego twórczości. Dobrze się stało, że dzięki życzliwości Tadeusza Bujnickiego, Czyżewskiego i Fałtynowicza teksty te nieoczekiwanie zbłądziły pod polonistyczne strzechy.

BIBLIOGRAFIA / BIBLIOGRAPHY

- Antologia poezji społecznej 1924-1933*. [To the Komitet Wydawniczy Czesław Miłosz, Zbigniew Folejewski]. Wilno: Koło Polonistów Sluchaczy Uniwersytetu Stefana Batorego, 1933.
- Balbus, Stanisław. “Pierwszy ruch jest śpiewanie” (O wierszu Miłosza – rozpoznanie wstępne).” In *Poznanwanie Miłosza*. Edited by Jerzy Kwiatkowski. Kraków: Wydawnictwo Literackie, 1985.
- Bernacki, Marek. *Tropienie Miłosza. Hermeneutyczna “bio-grafia” poety*. Kraków: Universitas, 2019.
- Bujnicki, Tadeusz. “Rękopisy Miłosza.” *Kroniki według Muzeum im. Marii Konopnickiej w Suwałkach*, no. 4(29) (2021): 20.
- Cudak Romuald, Wioletta Hajduk-Gawron, and Agnieszka Madeja, eds. „Na wrywki”: 100 cytatów z polskiej poezji i dramatu, które powinien znać także cudzoziemiec. Katowice: Wydawnictwo Uniwersytetu Śląskiego, 2018.
- Czechowicz, Józef. *Dzień jak co dzień*. Warszawa: f. hoesick, 1930.
- Daujotyte, Viktorija, and Mindaugas Kvietkauskas. *Litewskie konteksty Czesława Miłosza*. Translated by Joanna Tabor. Sejny: Pogranicze, 2014.
- Fałtynowicz, Zbigniew, “Dar.” *Kroniki według Muzeum im. Marii Konopnickiej w Suwałkach*, no. 4 (29) (2021): 20–25.
- Franaszek, Andrzej. *Miłosz. Biografia*. Kraków: Znak, 2011.
- Gorczyńska, Renata. “Miłosz wileński. Część druga.” *Kwartalnik Artystyczny*, no. 4 (2021): 75–94.
- Mann, Thomas. *Żarodziejska góra*. Vol. 2. Translated by Władysław Tatarkiewicz. Warszawa: Świat Książki, 1998.
- Miłosz, Czesław. “Bulion z gwoździ.” In *Miłosz, Przygody młodego umysłu. Publicystyka i proza 1931-1939*. Edited by Agnieszka Stawiarska and Aleksander Fiut. Kraków: Znak, 2003.

- . “Dom młodzińców.” In Miłosz. *Wiersze*. Vol. 1. Kraków: Znak, 2001.
- . “Kompozycja.” In Miłosz. *Wiersze*. Vol. 1. Kraków: Znak, 2001.
- . “Na śmierć młodego mężczyzny.” In Miłosz. *Wiersze*. Vol. 1. Kraków: Znak, 2001.
- . “Opowieść.” In Miłosz. *Wiersze*. Vol. 1. Kraków: Znak, 2001.
- . “Piękna jesteś ojczyzno moja, Europo...”. *Magazyn Świąteczny Gazety Wyborczej*, April 1–2, 2017: 33.
- . *Poemat o czasie zastygłym*. Wilno: Koło Polonistów Uniwersytetu Wileńskiego, 1933.
- . “Przeciwko nim.” In Miłosz. *Wiersze*. Vol. 1. Kraków: Znak, 2001.
- . “Rachunek znowu,” [rękopis]. *Kroniki według Muzeum im. Marii Konopnickiej w Suwałkach*, no. 4 (29) (2021): 23.
- . *Trzy zimy*. Wilno and Warszawa: Związek Zawodowy Literatów Polskich, 1936.
- . “Wam.” In Miłosz. *Wiersze*. Vol. 1. Kraków: Znak, 2001.
- . *Wiersze*. Vols. 1–5. Kraków: Znak, 2001–2009.
- . *Wiersze wszystkie*. Kraków: Znak, 2011.
- . *Wiersze wszystkie*. Kraków: Znak, 2015.
- . *Wiersze wszystkie*. Kraków: Znak, 2018.
- . *Wyprawa w dwudziestolecie*. Kraków: Wydawnictwo Literackie, 2000.
- . “Z dołu” [rękopis]. *Kroniki według Muzeum im. Marii Konopnickiej w Suwałkach*, no. 4 (29) (2021): 22.
- . “Z pełną świadomością...” (manuscript). *Kroniki według Muzeum im. Marii Konopnickiej w Suwałkach*, no. 4 (29) (2021): 21.
- , and Jarosław Iwaszkiewicz. *Portret podwójny: wykonany z listów, wierszy, zapisków intymnych, wywiadów i publikacji*. Edited by Barbara Toruńczyk and Robert Papieski. Warszawa: Fundacja Zeszytów Literackich, 2011.
- Wyka, Marta. *Miłosz i rówieśnicy: domknięcie formacji*. Kraków: Universitas, 2013.
- Zaleski, Marek. *Przygoda drugiej awangardy*. Wrocław: Zakład Narodowy im. Ossolińskich, 2000.
- Zawada, Andrzej. *Miłosz*. Wrocław: Wydawnictwo Dolnośląskie, 1996.

ABSTRAKT / ABSTRACT

Marek BERNACKI – „Czas przyszedł jedną ogarnie nas nazwą: / Niedoskonałych i już zapomnianych”. O nieznanym i niewiedzionym wierszach Czesława Miłosza

DOI 10.12887/36-2023-3-143-15

W artykule autor dokonuje analityczno-interpretacyjnego oglądu trzech nieznanymi dotąd młodzińskich wierszy Czesława Miłosza: *Rachunek znowu*, *Z dołu* i „*Z pełną świadomością...*”. Porównuje je z innymi utworami z tego okresu, które złożyły się na debiutancki tom Miłosza *Poemat o czasie zastygłym*

(1933). Omówione w artykule juvenilne teksty wileńskiego poety powstały w latach 1930-1931, następnie zostały zdeponowane w prywatnym archiwum przyjaciela z lat młodości Teodora Bujnickiego. Wiersze upublicznił 30 czerwca 2021 roku Tadeusz Bujnicki, syn Teodora. Po raz pierwszy zostały opublikowane (w formie fotokopii zachowanych rękopisów) w kwartalniku „Kroniki” redagowanym przez Zbigniewa Fałtynowicza.

Słowa kluczowe: juvenilna twórczość Czesława Miłosza, grupa poetycka Żagary, Druga Awangarda, język poetycki Czesława Miłosza, literackie Wilno

Kontakt: Katedra Polonistyki, Wydział Humanistyczno-Społeczny, Uniwersytet Bielsko-Bialski, ul. Willowa 2, 43-309 Bielsko-Biała
E-mail: mbernacki@ubb.edu.pl
ORCID: 0000-0001-8986-0733

Marek BERNACKI, “Time future will embrace us with one name / That of the imperfect and the already forgotten.” On the Unknown Juvenile Poems by Czesław Miłosz
DOI 10.12887/36-2023-3-143-15

In the article, the author provides an analytical and interpretative review of three hitherto unknown juvenile poems by Czesław Miłosz: *Rachunek znowu* (Account again), *Z dołu* (From the Bottom), “*Z pełną świadomością...*” (“With Full Consciousness...”), comparing them to other works from that period included in Miłosz’s debut volume *Poemat o czasie zastygłym* (A Poem on Frozen Time) (1933). The works in question were written in the years 1930-1931, deposited by the poet in the private archive of his friend Teodor Bujnicki, and discovered and made public by Teodor’s son, Tadeusz Bujnicki, only in 2021. The discussed poems were first published (as photocopies of the manuscripts) in the quarterly *Kroniki* edited by Zbigniew Fałtynowicz.

Keywords: early works of Czesław Miłosz, the Żagary poetic group, the Second Vanguard (Druga Awangarda), poetic language of Czesław Miłosz, literary Vilnius

Contact: Katedra Polonistyki, Wydział Humanistyczno-Społeczny, Uniwersytet Bielsko-Bialski, ul. Willowa 2, 43-309 Bielsko-Biała, Poland
E-mail: mbernacki@ubb.edu.pl
ORCID: 0000-0001-8986-0733

Robert KUBLIKOWSKI

FILOZOFIA SZTUCZNEJ INTELIGENCJI

Książka pod tytułem *Sztuczna inteligencja. Jej natura i przyszłość*¹ autorstwa Margaret A. Boden – opublikowana pierwotnie w Oxford University Press – dobrze wpisuje się w utrzymujące się od kilku dekad zainteresowanie tą problematyką².

Autorka monografii, o czym dowiadujemy się między innymi z informacji na okładce, jest profesorem nauk kognitywnych na Wydziale Informatyki Uniwersytetu w Sussex w Wielkiej Brytanii.

Polski, porządny przekład – wydany w serii wydawniczej „Krótkie Wprowadzenie” – został wykonany przez adiunkta Tomasza Sieczkowskiego, pracownika Wydziału Filozoficzno-Historycznego Uniwersytetu Łódzkiego. O naukową redakcję zadbał Piotr Fulmański, adiunkt na tamtejszym Wydziale Matematyki i Informatyki. Odredakcyjne przypisy objaśniają rzetelnie trudne fragmenty monografii.

W rozdziale pierwszym, „Czym jest sztuczna inteligencja?” (zob. s. 13-32), Boden wprowadza w tytułową problematykę. Zdaniem autorki sztuczna inteligencja (ang. artificial intelligence, AI), wykorzystuje różne techniki, których celem jest

to, aby komputer wykonywał takie same czynności (rozumowanie i temu podobne), jak umysł ludzki. AI ma dwa cele główne: (1) technologiczny (praktyczny), kładący akcent na użyteczność komputerów i (2) naukowy (teoretyczny), polegający na zastosowaniu pojęć i modeli w taki sposób, żeby za pomocą AI można było odpowiedzieć na pytania dotyczące organizmów żywych (między innymi wykorzystywany przez biologów model „sztucznego życia”, ang. A-life), a szczególnie odnoszące się do ludzi (zob. s. 13-15). To drugie zadanie jest mniej techniczne, a bardziej antropologiczne. Przy użyciu komputerów – ze względu na występowanie analogii między systemami informatycznymi a organizmami – można bowiem lepiej zrozumieć na przykład ludzkie umysły, ich naturę, strukturę, funkcje czy metody działania. Fundamentalnym pytaniem pozostaje, jakie są podobieństwa i różnice między inteligencją naturalną a sztuczną?

AI jest związana nie tyle z maszyną fizyczną (komputerem), ile z jakoś zdefiniowaną maszyną wirtualną, będącą symulacją analizowanego systemu. Maszyna wirtualna ma wykonywać procesy dokładnie w taki sposób, jak to się dokonuje w oryginalnym systemie fizycznym. Składa się ona ze wzorców działania – przetwarzania informacji (zob. s. 15-17).

Wyróżnia się pięć głównych typów AI: (1) klasyczna (symboliczna) AI (ang.

¹ Margaret A. Boden, *Sztuczna inteligencja. Jej natura i przyszłość*, tłum. T. Sieczkowski, red. P. Fulmański, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2020, ss. 208.

² Zob. Institute for Ethics in AI, www.oxford-aiethics.ox.ac.uk.

good old-fashioned AI, GOFAI). GOFAI zapoczątkował Alan Turing, wykazując, że wszystkie możliwe obliczenia mogą być zrealizowane przez system matematyczny, współcześnie określany jako uniwersalna maszyna Turinga. Przekonanie co do możliwości uzyskania AI wsparli neurolog i psychiatra Warren McCulloch oraz matematyk Walter Pitts, łącząc podejście Turinga z teorią synaps nerwowych autorstwa Charlesa Sherringtona i rachunkiem zdań Bertranda Russella. Arthur Samuel z kolei napisał program grający w szachy, który sam nauczył się wygrywać z jego twórcą. Natomiast Maszyna Teorii Logicznej (ang. logic theory machine), będąca w stanie udowadniać twierdzenia logiczne (Allen Newell, Herbert Simon), została prześcignięta przez Maszynę Rozstrzygającą Problemy (ang. general problem solver – Newell, John Shaw, Simon). Kolejnymi typami AI są: (2) sztuczne sieci neuronowe (koneksjonizm), (3) automaty komórkowe, (4) programowanie ewolucyjne i (5) systemy dynamiczne (zob. s. 18-32). Te typy AI autorka scharakteryzowała dokładniej w dalszej części swojej monografii.

Rozdział drugi, „Ogólna inteligencja jako Święty Graal” (zob. s. 33-69), dotyczy szansy utworzenia sztucznej inteligencji ogólnego przeznaczenia, czyli tak zwanej ogólnej sztucznej inteligencji (ang. artificial general intelligence, AGI). Gdyby udało się uzyskać AGI, to komputery byłyby w stanie funkcjonować nie tyle na mocy konkretnego, s z c z e g ó ł o w e g o oprogramowania dedykowanego do realizacji s z c z e g ó ł o w e g o celu, ile na podstawie możliwości percypowania, używania języka, rozumowania, kreatywności i tym podobnych w zastosowaniu do szerokiego, o g ó l n e g o spektrum problemów. Problemy mają być obliczalne efektywnie, czyli im mniej jest obliczeń, tym lepiej. Umożliwiają to następujące metody (strategie): (1) wyszukiwanie heurystyczne (akcent pada na takie zdefiniowanie

programu, aby był on ukierunkowany na konkretne aspekty przestrzeni przeszukiwania), (2) planowanie (na odpowiednim poziomie abstrakcji może ono skutkować przycięciem – zredukowaniem drzewa reprezentującego elementy sprawdzone oraz te wymagające sprawdzenia), (3) upraszczanie matematyczne (między innymi milcząco uznawane jest założenie zaniebujące rolę emocji w ludzkim działaniu) i (4) reprezentacja wiedzy w określonym języku programowania.

Reprezentacja wiedzy stosuje zbiór reguł JESLI-TO: JEŚLI dany Warunek jest spełniony, TO podejmij to Działanie.

Reprezentacja wiedzy może dotyczyć indywidualnych terminów (pojęć), a nie całych dziedzin wiedzy. Zdefiniowana hierarchiczna struktura (baza) danych to tak zwana r a m a. Problem ramy polega na tym, że programy AI nie mają ludzkiej zdolności do uchwycenia, rozumienia i s t o t n o ś c i, czyli tego, co jest ważne w konkretnej sytuacji. Problemu ramy można by uniknąć, gdyby były znane wszelkie możliwe k o n s e k w e n c j e każdego możliwego działania.

Reprezentacja wiedzy wymaga zdefiniowania stereotypowych sekwencji działań (na przykład określenie, co należy do „procedury” układania dziecka do snu).

W reprezentacji wiedzy akcentuje się tak zwane s i e c i s e m a n t y c z n e – terminy (pojęcia) pozostają w relacjach semantycznych, na przykład synonimiczności, antonimiczności, podporządkowania. Programy przeszukujące sieć internetową – silniki wyszukiwania – generalnie ujmując, programy NLP (ang. natural language processing) mogą znajdować związki między słowami czy tekstami, jednakże brak im r o z u m i e n i a tego, co przeszukują. Sieci semantyczne wiążą się również z poznaniem rozproszonym, czyli niehierarchizowanym (holizm semantyczny).

Reprezentacja wiedzy – czy szerzej ujmując, AGI na poziomie człowieka – musi

uwzględniać także uczenie maszynowe (ang. machine learning). Wyróżnia się uczenie: (1) nadzorowane, (2) nienadzorowane i (3) przez wzmocnienie. W uczeniu nadzorowanym programista szkoli system przez zdefiniowanie – dla ustalonych danych wyjściowych – zbioru oczekiwanych rezultatów. System uczący się formułuje hipotezy dotyczące stosownych własności i odpowiednio – w zależności od szczegółowej informacji zwrotnej o błędach – koryguje hipotezę wyjściową. W uczeniu nienadzorowanym użytkownik nie określa oczekiwanych rezultatów i nie otrzymuje komunikatów o błędach. Przyjmowana jest zaś reguła, że własności współwystępujące będą współwystępowały również w przyszłości. Natomiast uczenie przez wzmocnienie wykorzystuje schemat nagród (za sukces) i kar (za porażkę): system – na podstawie komunikatów zwrotnych – jest informowany, że jego dana reakcja była poprawna lub niepoprawna. Wyniki są aktualizowane po pojedynczej decyzji lub sekwencji decyzji.

W rozdziale trzecim, „Język, kreatywność, emocje” (zob. s. 71-91), została zaakcentowana rola NLP w tłumaczeniu maszynowym i trudności związane z analizami, zarówno syntaktyczną, jak i semantyczną. Niuanse struktury wypowiedzi powodują zmianę sensu i referencji. Oprócz tego w użyciu języka ważny jest również aspekt pragmatyczny: adekwatność (relevancja) użycia wyrażenia w danej sytuacji i kontekst użycia. Kontekst jest tworzony między innymi przez (1) wyrazy bliskoznaczne, (2) synonimy, (3) antonimy, (4) wyrazy, których zbiory (zakresy) są w relacji zawierania się w jakieś klasie, a także (5) wyrazy, których desygnaty są w relacji należenia do jakiejś klasy. Kontekstualne są dla siebie również (6) wyrazy, których desygnaty są w relacji część–całość. Częstość współwystępowania słów jest ważna przy wyszukiwaniu (eksplorowaniu) informacji.

Pojęcia związane z AI są użyteczne w wyjaśnianiu i porządkowaniu (typologizowaniu) ludzkiej kreatywności. Wyróżnia się: (1) kreatywność kombinacyjną – polegającą na tym, że znane idee są łączone w dotychczas nieznaną sposobą, (2) kreatywność eksploracyjną – bazującą na kulturowo znanych sposobach (stylach, regułach) myślenia i ich niedużych zmianach, (3) kreatywność transformacyjną – dopuszczającą duży zmianę reguł w celu uzyskania radykalnej nowości.

Dodowartościowania emocji w AI przyczynili się Marvin Minsky i Aaron Sloman, którzy postulowali, aby traktować umysł jako całość (intelekt-racjonalność, wola-decyzja, emocje).

Rozdział czwarty, zatytułowany „Sztuczne sieci neuronowe” (zob. s. 93-116), scharakteryzował je jako układy złożone z licznych, wzajemnie połączonych jednostek (elementów) przetwarzających. Każda z nich wykonuje tylko jeden rodzaj obliczeń. Jedną ze sztucznych sieci neuronowych jest przetwarzanie równoległe rozproszone (ang. parallel distributed processing, PDP). Zaletą PDP jest: (1) umiejętność uczenia się wzorców i związków między wzorcami za pośrednictwem przykładów, a nie przez bezpośrednie programowanie, (2) tolerancyjność względem nieprecyzyjnych informacji, (3) możliwość rozpoznawania niekompletnych lub częściowo uszkodzonych wzorców, (4) odporność na zniszczenia. Jeżeli uszkodzenia są coraz większe, to działanie PDP pogarsza się stopniowo (zob. s. 93-101).

System – w tak zwanym uczeniu głębokim (ang. deep learning) – przyswaja sobie strukturę głęboką, a nie tylko wzorce powierzchowne występujące w jakiejś dziedzinie (zob. s. 102-116). Ta problematyka łączy się z dyskutowanymi w filozofii języka i nauki – esencjalizmem, teorią rodzajów natu-

ralnych wraz z teorią terminów naturalno-rodzajowych, a także z teorią definicji ostensywnych i realnych (istotowych) oraz problemem tak zwanej *ramy*.

W rozdziale piątym, „Roboty i sztuczne życie” (zob. s. 117-134), autorka podkreśla, że roboty klasyczne z AI potrafiły naśladować *świadość i ałańia* człowieka. Nie były jednak w stanie reagować natychmiast na zmiany w otoczeniu. Natomiast roboty współczesne – tak zwane roboty *usytuowane* – są oparte na *reakcji na sytuację*, na zachowaniu adaptacyjnym. Zdolność do odruchów sensomotorycznych jest zasadniczo zapewniona dzięki anatomii robota (jego czujnikom i połączeniom sensomotorycznym), a nie jego oprogramowaniu.

Roboty – zwane też automatami komórkowymi – są systemami poszczególnych jednostek, z których każda przybiera jeden ze skończonej ilości stanów odpowiednio do prostych reguł zależnych od bieżącego stanu sąsiadujących jednostek składających się na automat (twórcą automatów komórkowych był John von Neumann).

Roboty, w których implementowane jest programowanie ewolucyjne (możemy to prześledzić w pracach na przykład Norberta Wienera), mogą się zmieniać (rewidować i korygować) dzięki algorytmom genetycznym inspirowanym przez biologię ewolucyjną.

Rozdział szósty, pod tytułem „Ale czy to naprawdę jest inteligencja?” (zob. s. 135-160), zawiera fundamentalne pytania na temat AI: Czy przyszłe systemy AGI będą dysponowały prawdziwą świadomością, inteligencją, rozumieniem i kreatywnością? Czy będą one posiadały status moralny i możliwość wolnego wyboru? Próbowano udzielić na nie odpowiedzi (na przykład test Turinga czy myślowy eksperyment Johna Searle’a o tak zwanym chińskim pokoju obrosły dyskusjami), lecz nie dostarczyły one bezspornych stanowisk.

W ostatnim rozdziale (siódmym), „Osobliwość” (zob. s. 161-183), zasygnalizowano dylematy moralne związane ze sztuczną inteligencją. Rozwój technologiczny AI zastępującej ludzi może zwiększyć bezrobocie, ale też skutkować pojawieniem się nowych zawodów i miejsc pracy. Wątpliwości etyczne budzi również konstruowanie robotów erotycznych (sztucznych partnerów seksualnych) czy robotów do opieki nad ludźmi niepełnosprawnymi. Czy sztuczna empatia jest moralnie dobra i akceptowalna? Czy budowanie i używanie takich robotów jest etycznie uzasadnione? Podobny dylemat dotyczy robotów militarnych, skonstruowanych do likwidacji celów wskazanych przez obsługujący je personel, a nawet wybranych samodzielnie przez robota. Poszukiwane są kodeksy postępowania dla robotów, mające rozwinąć „kompetencje moralne” AI. Przykładem są „trzy prawa robotyki” Issaca Asimova: (1) robot nie powinien skrzywdzić człowieka, (2) robot musi wypełniać ludzkie polecenia, (3) robot powinien dbać o własne przetrwanie. W przypadku konfliktu interesów pierwszeństwo ma prawo pierwsze. Idealem byłoby utworzenie systemu obliczeniowego zdolnego do przeprowadzania rozumowań i dyskusji moralnych.

Książka została zaopatrzona także w obszerną „Bibliografię” (zob. s. 185-199) i rozbudowany „Indeks osobowo-rzeczowy” (zob. s. 201-208).

Podsumowując, styl pisarski autorki monografii mógłby być niekiedy precyzyjniejszy (na przykład w początkowym rozdziale badaczka nie wskazuje dokładnie, które typy AI są reprezentowane przez których twórców). Jednak nie jest to częste. Recenzowana książka zasadniczo jest wartościowym – merytorycznym, erudycyjnym, a jednocześnie wyważonym i zwartym – wprowadzeniem w zagadnienia związane z jednej strony z informatyką, a z drugiej – z filozofią nauki i techniki, a dokładniej z filozofią (ontologią,

semiotyką, metodologią, epistemologią czy etyką) sztucznej inteligencji. Dyscypliny te mogą być dla siebie nawzajem inspiracją do badań uwzględniających nowe konteksty i wyzwania teoretyczne oraz praktyczne.

Kontakt: Katedra Metodologii Nauk, Instytut Filozofii, Wydział Filozofii, Katolicki Uniwersytet Lubelski Jana Pawła II, Al. Raclawickie 14, 20-950 Lublin
E-mail: robert.kublikowski@kul.pl
ORCID: 0000-0002-5199-3941

PROPOZYCJE „ETHOSU”

Jobst Landgrebe, Barry Smith, *Why Machines Will Never Rule the World: Artificial Intelligence without Fear*, Routledge, New York – London 2022, ss. 341.

Rozwój sztucznej inteligencji (AI) jest tematem zaprzatającym umysły naukowców z różnych dyscyplin, polityków, przedsiębiorców i zwykłych ludzi, a postawy wobec tej kwestii wahają się od bezkrytycznego entuzjazmu do autentycznego strachu przed buntem inteligentnych robotów, które mogą uznać nas, ludzi, za niższy gatunek, podporządkowując nas, a ostatecznie – eliminując. Książka Jobsta Landgrebe’a, naukowca i przedsiębiorcy z filozoficznym wykształceniem, i Barry’ego Smitha, wybitnego filozofa, jest odpowiedzią na te obawy: silna (ang. general) sztuczna inteligencja jest matematycznie – nie zaś jedynie technicznie czy fizykalnie – niemożliwa. „Spokojnie. Maszyny nigdy nie będą rządziły światem” (s. x; tłum. fragmentów książki – A.L.K.) – piszą autorzy książki. Definiują oni sztuczną inteligencję ogólną (AGI) jako tę, która prezentuje poziom wyższy od inteligencji ludzkiej, lub co najmniej jej równy, i jest zdolna do radzenia sobie z problemami pojawiającymi się w otaczającym człowieka świecie w stopniu adekwatności co najmniej podobnym do tego, z jakim radzi sobie z nimi człowiek. Nie negują oczywiście postępu w budowaniu sztucznych inteligencji przeznaczonych do konkretnych zadań, kwestionują natomiast nieograniczony postęp w symulacji działania człowieka. Argument na rzecz ich tezy, w nieco uproszczonej formie, przedstawia się następująco: (1) zbudowanie AGI wymaga napisania oprogramowania, które naśladowałoby ludzki system neuropoznawczy (autorzy rozważają i ostatecznie odrzucają alternatywne sposoby wykreowania AGI); (2) by napisać takie oprogramowanie, należałoby zbudować matematyczny model tego systemu, umożliwiający przewidywanie jego zachowania; (3) nie jesteśmy w stanie budować matematycznych modeli dynamicznych systemów złożonych, choć systemy te podlegają prawom fizyki; (4) ludzki system neuropoznawczy jest dynamicznym systemem złożonym, a wobec tego nie jesteśmy w stanie zbudować wymaganego matematycznego modelu tego systemu i dlatego nie możemy napisać oprogramowania, które naśladowałoby ludzki system neuropoznawczy. Ergo: stworzenie AGI jest niemożliwe z powodu ograniczeń płynących z matematyki. Pozyskiwanie pieniędzy na badania nad AGI zostało więc oparte na

falszywych obietnicach, a strach przed maszynami rządzącymi światem jest nieuzasadniony.

W pierwszej części książki Landgrebe i Smith w kolejnych rozdziałach prowadzą nas przez analizy ludzkiego umysłu i problemu relacji między umysłem a ciałem, ludzką i maszynową inteligencją, zagadnienia natury ludzkiego języka i komunikacji oraz ludzkich zachowań społecznych i moralnych. W dwóch rozdziałach części drugiej autorzy poddają refleksji naturę systemów złożonych i rozmaite podejścia do ich matematycznego modelowania, wskazując na wsobne ograniczenia tego procesu. Im bardziej złożony jest system, tym mniejsze są szanse na jego adekwatne modelowanie i tworzenie dla niego AI. Część trzecia poświęcona została możliwościom sztucznej inteligencji i jej ograniczeniom. Tytuły kolejnych rozdziałów stanowią zarazem tezę, że maszyny nie osiągną oczekiwanego poziomu inteligencji, nie opanują ludzkiego języka i nie opanują interakcji społecznych w stopniu wyższym czy choćby równym temu, który jest dostępny człowiekowi. Możliwe jest natomiast matematyczne modelowanie częściowe – i właśnie tam znajduje zastosowanie sztuczna inteligencja.

W dyskusjach autorzy sięgają do przykładów z medycyny, inżynierii, ekonomii, psychologii, odpowiadając na obietnice i obawy wielu współczesnych myślicieli, w tym wiodących teoretyków transhumanizmu: Nicka Bostroma i Raya Kurzweila czy Elona Muska. Rozważania Landgrebe'a i Smitha pozostają pod wyraźnym wpływem fenomenologii, w tym Edmunda Husserla (z okresu *Doświadczeń filozoficznych*), Adolfa Reinacha i Maxa Schelera, czego autorzy bynajmniej nie ukrywają, powołując się na tych myślicieli, zwłaszcza w odniesieniu do analiz natury człowieka, jego poznania i woli, a jednocześnie pokazując niewystarczalność rozmaitych podejść filozofii analitycznej do uchwycenia natury naszego świata. Autorzy deklarują, że nadrzędną ideą, która spaja wszystkie części książki, jest uznanie istnienia nieredukowalnych do siebie rzeczywistości: fizycznej, biologicznej, społecznej i mentalnej, oraz przyjęcie realistycznej filozofii dotyczącej świata dostępnego w naszym codziennym doświadczeniu (por. s. 3n.). To podejście – wraz z odwołaniem do Maxa Schelera koncepcji inteligencji praktycznej – pozwala autorom odróżnić inteligencję pierwotną od obiektywizującej (ang. objectifying) czy uprzedmiotowującej. Ta pierwsza przysługuje bytom ludzkim i pozaludzkim (choćby zwierzętom) i nie jest rezultatem uczenia się, ale raczej umożliwia uczenie się poprzez adaptację do nowych sytuacji oraz rozwinięcie pierwotnego zachowania kulturotwórczego. Ta druga przysługuje ludziom i umożliwia kumulatywną ewolucję kultury: „Inteligencja obiektywizująca pozwala homo sapiens zdystansować się od swojego środowiska w sposób, który umożliwia widzenie siebie, innych ludzi i elementów tego środowiska (biologicznych i niebiologicznych) jako przedmiotów, z których każdy ma własną trajektorię i własny zbiór właściwości i mocy kauzalnych” (s. 46). Inteligencja ta obejmuje zdolność ujmowania świata poprzez kategorie ogólne, intencjonalność, rozpoznawanie przyczyn i celów, panowanie nad koniecznościami biologicznymi, zdolność decydowania (czyli wyboru i niejako ustanawiania siebie źródłem działania) oraz wyobrażania sobie i planowanie długoterminowe. To wszystko musiałaby posiadać AGI, a skoro nie da się tego modelować matematycznie, to inteligencja sztucznej inteligencji – jakkolwiek by ją rozumieć, co autorzy starannie dyskutują – nie

przekroczy inteligencji ludzkiej ani nawet jej nie dorówna. Ludzki mózg nie jest komputerem i żadna znana matematyka nie może go całkowicie adekwatnie opisać. Landgrebe i Smith ostatecznie głoszą zatem wyjątkowość ludzkiego bytu. Sztuczna inteligencja nas nie zastąpi. Paradoksalnie, dyskusja nad AGI włącza się pośrednio w dyskusję nad naturą człowieka – nad tym, czy jesteśmy jedynie zwierzętami, czy też jesteśmy „inaczej i wyżej”, by przywołać frazę ukutą przez Tadeusza Stycznia, reprezentanta lubelskiej szkoły filozoficznej. A jeśli faktycznie status bytowy człowieka sytuuje nas wyżej, to jaka jest nasza właściwa relacja do bytów pozaludzkich, jakie mamy wobec nich prawa i obowiązki?

Argumentacja Landgrebe’a i Smitha odwołuje się do filozofii (różnych tradycji), matematyki, informatyki, językoznawstwa, psychologii, antropologii kulturowej, socjologii, fizyki i biologii. W tym sensie są to rozważania naprawdę interdyscyplinarne, co pokazuje naturę problemu sztucznej inteligencji – nie jest to bowiem problem czysto techniczny. Owa interdyscyplinarność nie ułatwia lektury książki. Zdając sobie z tego sprawę, autorzy wyposażyli ją w słownik kluczowych pojęć, indeks i aneks z matematycznym ujęciem turbulencji. Wskazują też fragmenty, które czytelnik może bez szkody dla zrozumienia argumentacji pominąć. Bibliografia zawiera imponującą liczbę pozycji i sama może stanowić źródło dla badaczy AI.

Książka wpisuje się w literaturę kwestionującą nieograniczone możliwości sztucznej inteligencji. Jednym z pierwszych autorów prezentujących takie stanowisko był Hubert Dreyfus, który w roku 1972 opublikował książkę *What Computers Can't Do: A Critique of Artificial Reason* (Harper and Row, New York-Evanston-San Francisco-London). Do listy można dopisać Rogera Penrose’a *Shadows of the Mind: A Search for the Missing Science of Consciousness* (Oxford University Press, Oxford 1994) czy Roberta J. Marksa II, *Non-computable You: What You Do That Artificial Intelligence Never Will* (Discovery Institute, Seattle 2022). Nawet jeśli Landgrebe i Smith mają rację – a rodzi się przecież szereg pytań, choćby o nową matematykę, o możliwości komputerów kwantowych czy znalezienie strategii budowania AGI innej niż symulacja ludzkiego umysłu – nie możemy spocząć w spokoju ducha. Jeśli nawet AGI jest niemożliwa, wciąż rozwija się przecież „zwykła” sztuczna inteligencja. „Mimo poważnych ograniczeń, które opisaliśmy w tej książce – twierdzą autorzy – odkrycie coraz bardziej pomysłowych i potężnych zastosowań wąskiej woli sztucznej inteligencji w nadchodzących dekadach w ogromnym stopniu pogłębi i wzmocni „technosferę” (s. 301), zmieniając świat naszego codziennego doświadczenia. Wyzwaniem – jak piszą – nie jest zapobieganie zastąpieniu ludzkości przez superinteligencję czy tworzenie „etycznej” AI, ale przewidywanie konsekwencji pojawiania się „wąskiej” sztucznej inteligencji w różnych sektorach życia i reagowanie na te konsekwencje. Można złośliwie twierdzić, że AI już do pewnego stopnia rządzi światem, bo wiele naszych decyzji sterowanych jest informacjami gromadzonymi przez sztuczną inteligencję, nasze czynności są zapośredniczone przez urządzenia wyposażone w AI, a sztuczna inteligencja staje się uczestnikiem naszego życia społecznego (na przykład w postaci empatycznych robotów). Obecność AI zmienia nas samych, nasze relacje z innymi ludźmi i nasze rozumienie świata. Pytania o to, gdzie, dlaczego

i pod jakimi warunkami powinniśmy wprowadzać sztuczną inteligencję, stają się coraz bardziej istotne. Warto zatem przeczytać książkę Landgrebe’a i Smitha, by z jednej strony lepiej zrozumieć, o co należy, a o co nie należy się martwić w związku z rozwojem sztucznej inteligencji, a z drugiej strony, by uświadomić sobie, kim jesteśmy i co jest stawką w rozwijaniu AI. Co więcej, książka jest znakomitym dowodem na to, że bez filozofii nie udaje się podjąć palących problemów współczesności (o czym zdają się zapominać adwokaci stawiania na kształcenie praktyczne, które – jak twierdzą – jest niezbędne do uzyskiwania przewagi na rynku pracy).

A.L.K.

Mirosława CHUDA

OCZEKIWANIE

Rozważania o sztucznej inteligencji (SI) w sposób nieunikniony kierują myśl ku przyszłości. Zarówno w prognozowaniu naukowym, w debacie publicznej, jak i w świadomości użytkowników urządzeń wykorzystujących SI obecna jest troska o konsekwencje postępu w tej dziedzinie. Z jednej strony wyrażane są obawy przed niekorzystnym wpływem wysoko rozwiniętych technologii tego rodzaju na człowieka i możliwościami zgubnego dla ludzi ich zastosowania, przed radykalnym przeobrażeniem znanego świata czy nawet podporządkowaniem ludzkości tworum techniki, z drugiej zaś – nadzieje, czy to na łatwiejsze funkcjonowanie w życiu codziennym i zawodowym, czy na przykład na większą skuteczność działań w zakresie ochrony zdrowia. Troska owa ma racjonalne podstawy, obawy niekiedy jednak wynikają z przesądów, lęku przed tym, co nieznanie i niezrozumiałe, jawiące się jako „wiedza tajemna”, a nadzieje z przesadnej ekscytacji osiągnięciami cywilizacyjnymi. U podstaw jednych i drugich leży zaś właściwa człowiekowi tendencja do skupiania uwagi na przyszłości, zakorzeniona w kondycji człowieka, w jego istnieniu w czasie. Pewnym bardzo subiektywnym aspektem owego „wychylenia ku przyszłości”, związanym z doświadczeniem przygodności ludzkiego bytu i nieprzewidywalności własnego losu, poświęćmy trochę uwagi na marginesie zawartych w tym tomie rozważań.

Święty Augustyn, który w *Wyznaniach* wyrażał przypuszczenie, że czas jest „rozciągłością samego umysłu”¹, uważał, że przyszłość jako „diedzina czasu” istnieje jedynie w duszy, jest pewnym stanem świadomości: „Właściwie nie należałoby mówić, że istnieją trzy dziedziny czasu – przeszłość, teraźniejszość i przyszłość. Może ściślej było by takie ujęcie, że istnieją następujące trzy dziedziny czasu: teraźniejszość rzeczy minionych, teraźniejszość rzeczy obecnych, teraźniejszość rzeczy przyszłych. [...] Teraźniejszością rzeczy przeszłych jest pamięć, teraźniejszością rzeczy doczesnych jest dostrzeganie, teraźniejszością rzeczy przyszłych – oczekiwanie”².

Abstrahując od kontekstu – biskupa z Hippony koncepcji czasu – należy przyznać, że kategoria oczekiwania dobrze charakteryzuje wspomnianą wyżej tendencję. Bez owego oczekiwania nie sposób funkcjonować, towarzyszy ono człowiekowi na każdym etapie jego egzystencji. Dzieci oczekują osiągnięcia dorosłości, wyzwolenia z ograniczeń narzucanych im przez opiekunów, ludzie młodzi spodziewają się urzeczywistnienia swoich planów i pragnień, starzy zaś, gdy życie zweryfikuje już ich dążenia i aspiracje, oczekują choćby chwilowego złagodzenia bólu, słonecznej pogody

¹ Augustyn, *Wyznania*, ks. XI, 26, tłum. Z. Kubiak, Instytut Wydawniczy Pax, Warszawa 1982, s. 237.

² Tamże, ks. XI, 20, s. 232.

czy po prostu następnego posiłku. Starości, którą Ryszard Przybylski nazwał „cełą śmiercią”³, nieuchronnie towarzyszy też oczekiwanie kresu.

Wybieganie myślą poza chwilę obecną poprzez wyobrażanie sobie tego, co może lub musi się zdarzyć, mentalne antycypowanie urzeczywistnienia pragnień czy ziszczenia się obaw, niekiedy jednak nadmiernie człowieka angażuje. Permanentne koncentrowanie się na tym, co dopiero ma nadejść, skutkuje wówczas ignorowaniem chwili obecnej. Pascal twierdził wręcz, że tego rodzaju zaniedbywanie teraźniejszości jest zjawiskiem powszechnym, właściwością człowieka. „Nie dbamy nigdy o czas obecny. Uprzedzamy przyszłość, jako zbyt wolno nadchodzącą, jak gdyby chcąc przyspieszyć jej bieg; albo przywołujemy przeszłość, aby ją zatrzymać jako zbyt skwapliwą. Tak nierozsądni jesteśmy, że błakamy się w czasach, które nie są nasze, nie myśląc o jedynym, który nam przynależy”⁴.

Być może leży w naturze ludzkiej pragnienie „czegoś więcej” niż to, co oferuje teraźniejszość. Naturalną rzeczą jest również dążenie do lepszego, bardziej wartościowego życia czy nawet do odegrania doniosłej roli w świecie – człowiek dojrzały winien jednak podejmować działania zmierzające do urzeczywistnienia tych pragnień, nawet jeśli jego zaangażowanie nie miałoby doprowadzić do osiągnięcia zamierzonych rezultatów w pełni. A przecież zdarza się, że ktoś żywi wielkie pragnienia, lecz biernie oczekuje, iż w nieokreślonej przyszłości zostaną one spełnione bez jego usilnych starań, na skutek jakiegoś zrządzenia losu. Do takiej postawy nawiązuje wiersz Reinera Marii Rilkego *Przypomnienie*:

A ty czekasz, ty czekasz na jedno,
co twe życie wzniesie nieskończenie,
na wielkie, niezwykle zdarzenie,
na kamieni nagle przebudzenie,
na głębie, co u nóg twych legną⁵.

„Ty” liryczne w utworze spodziewa się przełomowego wydarzenia o randze niemal kosmicznej, które w sposób zasadniczy odmieni jego życie, nadając mu niewyobrażalnie wysoką rangę. Oczekując na tę rewolucyjną zmianę, zwraca się ku swej przeszłości i rozpamiętuje straty – jego wspomnienia pozwalają przypuszczać, że prezentuje raczej postawę obserwatora niż kreatora. Nagle uświadamia sobie, że owa zmiana nie nadejdzie, że tylko powtarzać się będzie „Postać, Modlitwa i Trwoga”⁶ minionych lat.

Na niezwykle, wielkie zdarzenie, które zmieni jego życie, oczekuje również John Marcher, bohater opowiadania Henry’ego Jamesa *Bestia w dżungli*⁷, inaczej jednak niż adresat słów podmiotu lirycznego w przytoczonym fragmencie wiersza, przekonany

³ R. Przybylski, *W celi śmierci*, w: tenże, *Baśń zimowa. Esej o starości*, Wydawnictwo Siel, Warszawa 1998, s. 101.

⁴ B. Pascal, *Myśli*, 168 [21], tłum. T. Żeleński (Boy), Instytut Wydawniczy Pax, Warszawa 1977, s. 82.

⁵ R.M. Rilke, *Przypomnienie*, w: tenże, *Poezje*, tłum. M. Jastrun, Wydawnictwo Literackie, Kraków 1974, s. 53.

⁶ Tamże.

⁷ Zob. H. James, *Bestia w dżungli*, w: tenże, *Daisy Miller*, tłum. J. Olędzka, Czytelnik, Warszawa 1986, s. 189-237.

jest on, że będzie to coś strasznego, co pojawi się nagle i co w niepojęty sposób może go zniaździć. Kieruje nim przecucie, że prędzej czy później doświadczy swego rodzaju katastrofy, i przeświadczenie to kształtuje jego myślenie i determinuje postępowanie. Wierzy, że jego przeznaczeniem jest czekanie na owo wydarzenie, nie zaś spełnianie się w działaniu. Wyjaśnia to swojej dawnej znajomej: „Wie pani, to nie jest coś, co ja muszę zrobić, dokonać na świecie, by się wyróżnić czy wzbudzić podziw. [...] No, raczej muszę na to poczekać, spotkać się z tym twarzą w twarz”⁸. Ma świadomość, że będzie musiał ponieść konsekwencje tego spotkania, nad obawą przed nimi dominuje jednak swego rodzaju poczucie satysfakcji z przyjętej postawy – przeświadczenie, że oczekiwanie na katastrofę w jakiś sposób czyni go wyjątkowym. Jego przyjaciółka, May, która – powodowana uczuciem – decyduje się towarzyszyć mu w tym oczekiwaniu na nieznane, po pewnym czasie zdaje sobie sprawę, że z powodu swojej obsesji, mimo licznych zdolności, osiągnął on tak niewiele, że jego zaangażowanie w pracę w urzędzie państwowym, obowiązki towarzyskie i codzienne działania było pozorne, skazane na wyparcie ze świadomości. Stara się jednak chronić go przed tą wiedzą. Z czasem Marcher zaczyna podejrzewać, że wypatrywana bestia – jak określał spodziewane wydarzenie – nigdy nie wyskoczy z lasu i że to oczekiwanie było istotą jego losu. Katastrofa jednak następuje – bohater opowiadania „spogląda jej w oczy”, gdy nagle uświadamia sobie, że jego życie było „zamierzoną dokładnie pustką”⁹, że wpatrzony w to, co miało nadejść, przeoczył szansę na miłość, która nadałaby sens jego egzystencji. A stało się tak dlatego, że „czekanie samo w sobie jest pochłaniające”¹⁰.

Ludzka tendencja do nadmiernego koncentrowania się na przyszłości niekiedy bywa postrzegana jako ucieczka przed teraźniejszością. Pisał wszak Pascal: „Zazwyczaj bowiem teraźniejszość nas rani. Ukrywamy ją naszym oczom, ponieważ nam jest przykra; a jeśli jest miła, żalujemy że się nam wymyka. Staramy się podeprzeć ją przyszłością, i siliłyśmy się rozrządzać rzeczami, które nie są w naszej mocy, na czas, którego nie wiemy zgoła czy dożyjemy”¹¹. Teraźniejszość jest trudna, ponieważ reagowanie na wyzwania chwili obecnej wymaga odwagi – zawsze bowiem wiąże się z odpowiedzialnością i pewnym ryzykiem: podejmowane decyzje nierzadko okazują się błędne, a działania mogą ujawnić brak zdolności czy niekompetencję tego, kto je realizuje. Przyszłość kształtowana przez konsekwencje spełnionych czynów często bardzo różni się od tej oczekiwanej. Dopóki człowiek jedynie snuje wizje tego, co pragnie osiągnąć, nie musi liczyć się z własnymi ograniczeniami ani przeciwnościami, na które zwykle się napotyka, urzeczywistniając swe pragnienia. Zbyt wygórowane cele i obawa przed konfrontacją z rzeczywistością prowadzą do ciągłego odkładania ich realizacji na nieokreślony odległy czas.

Bohater innego opowiadania Jamesa¹², Theobald, pragnie stworzyć dzieło na miarę dawnych mistrzów, których cześć – obraz Madonny stanowiący „résumé wszystkich Madonn włoskiej szkoły”¹³. Ta idée fixe zawładnęła malarzem bez reszty, porzucał

⁸ Tamże, s. 197.

⁹ Tamże, s. 236.

¹⁰ Tamże, s. 206.

¹¹ P a s c a l, dz. cyt., s. 82.

¹² Zob. H. J a m e s, *Madonna przyszłości*, w: tenże, *Madonna przyszłości i inne opowiadania*, tłum. M. Skibniewska, PIW, Warszawa 1962, s. 6-63.

¹³ Tamże, s. 31.

„zwykle światowe dążenia jedne po drugich”¹⁴. Zamiast rozwijać swój talent, nieustannie, latami rozwijał w myślach koncepcję obrazu, by uczynić ją doskonałą. Nie próbował przelać swojej wizji na płótno. Znajomym, spodziewającym się ujrzeć efekty jego pracy, tłumaczył, że wielkie dzieła poprzedzać musi kontemplacja i tajemnica. „Sztuka ma dużo czasu. [...] Ona może czekać”¹⁵ – twierdził, jakby nie zdając sobie sprawy z tego, że jego własny czas jest ograniczony. Gdy życzliwy młody człowiek uświadomił mu, że musi wreszcie zdobyć się na czyn, bo jego modelka i muza już się zestarzała i zachowała tylko resztki dawnej urody, przyrzekł skończyć obraz. To, co było możliwe w zamyśle, w praktyce okazało się jednak nierealne. Przebudzenie z marzeń doprowadziło Theobalda do rozpaczliwej konstatacji: „Roztrwoiłem czas! Jestem bankrutem! [...] Wciąż czekałem i czekałem, żeby stać się godniejszym podjęcia dzieła... Strwoiłem życie na przygotowaniach”¹⁶.

Przyczyn i skutków ucieczki ku przyszłości można by wskazać znacznie więcej. Można by też przytoczyć jeszcze wiele literackich przykładów tej postawy, poczynawszy od emblematycznego dla poruszonego tu zagadnienia dramatu Samuela Becketta¹⁷. Większość z nich mówi o porażce: to, co oczekiwane, nigdy nie nadchodzi, a jeśli się zdarzy, to przynosi rozczarowanie, „odkładanie życia na później” nieuchronnie prowadzi do utraty szans na samospełnienie, a koncentrowanie się na tym, co dopiero ma się wydarzyć, nie pozwala cieszyć się chwilą obecną i docenić tego, ona oferuje.

Często można spotkać się z opinią, że problemy te są bolączką szczególnie naszych czasów, gdyż presja społeczna każe człowiekowi wytyczać sobie coraz bardziej ambitne zadania, często przekraczające jego możliwości, a propagowane w mediach wzorce napędzają aspiracje i kształtują przekonanie o konieczności osiągnięcia spektakularnego sukcesu zawodowego i finansowego czy wysokiej pozycji w sferze publicznej. Ponieważ nadmierne skupianie się na przyszłości znacząco obniża skuteczność działania, poszukiwane są środki zaradcze pozwalające przezwyciężyć tę tendencję. Jednym z takich remediów jest – popularna zwłaszcza na Zachodzie – praktyka uważności (ang. *mindfulness*). Jak pisze Jon Kabat-Zinn, założyciel Center for Mindfulness in Medicine, Health Care, and Society na uniwersytecie w Massachusetts, ma ona prowadzić do „przebywania w teraźniejszości – przebywania w niej w sposób przytomny, spokojny, z jasnym umysłem i troską”¹⁸, a ten stan świadomości wypracować można poprzez intencjonalne i nieocenające skupianie uwagi na tym, czego doświadcza się tu i teraz. Kabat-Zinn opracował inspirowane głównie buddyzmem techniki medytacji pozwalające „odzyskać” teraźniejszość – wśród technik tych zaleca między innymi „niedążenie”, czyli pozostawanie w „bezczasowości chwili obecnej”¹⁹, bez wybiegania myślą w przyszłość²⁰.

¹⁴ Tamże, s. 14.

¹⁵ Tamże, s. 15.

¹⁶ Tamże, s. 56n.

¹⁷ Zob. S. B e c k e t t, *Czekając na Godota*, tłum. i oprac. A. Libera, PIW, Warszawa 1985.

¹⁸ J. K a b a t - Z i n n, *Praktyka uważności dla początkujących*, tłum. J.P. Listwan, Wydawnictwo Czarna Owca, Warszawa 2014, s. 24.

¹⁹ Tamże, s. 138.

²⁰ Treningi *mindfulness* znalazły szerokie zastosowanie: nie tylko w różnych nurtach psychoterapii, ale także w edukacji czy w biznesie – przedsiębiorcy organizują je dla swoich pracowników, aby podnieść jakość ich funkcjonowania w firmie (zob. np. L. M a r s h, *Mindfulness robi wam wodę*

Ludzka orientacja temporalna na przyszłość w każdej epoce musiała rodzić problemy, tym zaś towarzyszyć musiały próby radzenia sobie z nimi. Wystarczy przypomnieć Horacjańskie „carpe diem”²¹, które od projektu mindfulness dzieli era. Rzymski poeta, zwolennik filozofii epikurejskiej i stoickiej, zalecał Leuconoe:

Bądź mądra, wino cedź, nadzieję, póki nie pryska,
powściągaaj. Nawet teraz, gdy mówimy, umyka
czas nienawistny. Jutra nie czekaj. Z dnia korzystaj²².

Przytoczone przykładowo dwa podejścia do funkcjonowania człowieka w czasie zakładają, że możliwe jest uwolnienie się od oczekiwania w prezentowanym tu rozumieniu i że skupienie się na doświadczaniu teraźniejszości wiedzie do szczęśliwego życia. Pascal natomiast zdecydowanie temu przeczy: „Niechaj każdy zbada swoje myśli, ujrzy, iż wszystkie zaprzątnięte są przeszłością i przyszłością. Nie myślimy prawie zupełnie o teraźniejszości; a jeśli myślimy, to jeno aby zaczerpnąć z niej treść do snucia przyszłości. Teraźniejszość nie jest nigdy naszym celem: przeszłość i teraźniejszość to nasze środki; jedynie przyszłość jest naszym celem. Tak więc, nie żyjemy nigdy, ale spodziewamy się żyć; gotujemy się wciąż do szczęścia, a, co za tem idzie, nie kosztujemy go nigdy”²³. Bohaterowie przywołanych wyżej opowiadań Jamesa do czasu uświadomienia sobie, że nie ma już na co czekać, są jednak na swój sposób szczęśliwi. To, co oczekiwane, jest bowiem potencjalnie obecne tu i teraz.

z mózgu, <https://krytykapolityczna.pl/gospodarka/mindfulness-korporacje-prawa-pracownicze/>). Praktykowanie uważności miałoby pomóc w radzeniu sobie ze stresem. w przezwyciężaniu sytuacji kryzysowych, w leczeniu chorób psychicznych i somatycznych, miałaby uzdrowić relacje z innymi, ze światem i z samym sobą, pomóc odnosić sukcesy w różnych dziedzinach, nawet w sporcie. Jako swoiste panaceum praktyka ta musi jednak budzić wątpliwości, zwłaszcza że niesie z sobą ryzyko niepożądanych skutków dla ludzkiej psychiki (zob. np. L. S a n t a m a r i a, *Francuskie szkoły zalewane przez praktykę mindfulness. „Może wykształcać w dzieciach uległość i zanik krytycznego myślenia”*, <https://pl.aleteia.org/2022/02/10/francuskie-szkoly-zalewane-przez-praktyke-mindfulness-moze-wykszaltcac-w-dzieciach-uleglosc-i-zanik-krytycznego-myslenia/>).

²¹ H o r a c y, *Pieśni*, ks. I, pieśń XI, tłum. S. Gołębiowski, oprac. G. Pianko, L. Winniczuk, Ludowa Spółdzielnia Wydawnicza, Warszawa 1972, s. 64.

²² Tamże, s. 65.

²³ P a s c a l, dz. cyt., s. 82.

NOTY O AUTORACH

Marek B e r n a c k i, doktor habilitowany, profesor UBB, literaturoznawca. Studia z zakresu filologii polskiej na Uniwersytecie Jagiellońskim.

Od 2003 r. pracownik Akademii Techniczno-Humanistycznej w Bielsku-Białej, obecnie profesor uczelni w Katedrze Polonistyki na Wydziale Humanistyczno-Społecznym Uniwersytetu Bielsko-Bialskiego.

Redaktor naczelny czasopisma naukowego „Media i Społeczeństwo”.

Członek Komisji Historycznoliterackiej PAN (Oddział w Katowicach), członek założyciel Towarzystwa Filozoficznoliterackiego „Hermeneutica”.

Główne obszary zainteresowań badawczych: literatura polska XIX-XXI w., miłoszologia, sztuka interpretacji, hermeneutyka.

Najważniejsze publikacje książkowe: *Leksykon powieści polskich XX wieku* (współred., 2002); *Hermeneutyka fenomenu istnienia. Studia o polskiej literaturze współczesnej* (Vincenz, Miłosz, Wojtyła, Herbert, Szymborska) (2010); *Świat interpretować – konieczne zadanie. Studia o literaturze polskiej przełomu XX i XXI wieku* (2014); *Miłosz. Dyskursy* (współred., 2016); *Miłosz. Spotkania. Studia i rozprawy miłoszologiczne* (2018); *Tropienie Miłosza. Hermeneutyczna „bio-grafia” Poety* (2019); *Peryferie Miłosza. Nieznane konteksty, głosy, nowe rozpoznania* (red., 2020); *Interpretując: Jestem. Artykuły i szkice krytycznoliterackie (2017-2023)* (2023).

Ks. Marcin F e r d y n u s, doktor, filozof, teolog. Studia z zakresu teologii na Uniwersytecie Szczecińskim oraz na Uniwersytecie im. Adama Mickiewicza w Poznaniu, z zakresu filozofii na Katolickim Uniwersytecie Lubelskim Jana Pawła II.

Od 2017 do 2018 r. wykładowca w Wyższej Szkole Integracji Europejskiej w Szczecinie. Od 2016 do 2019 r. prefekt Katolickiego Liceum Ogólnokształcącego w Szczecinie. Od 2017 do 2019 r. dyrektor do spraw formacji w Centrum Edukacyjnym Archidiecezji Szczecińsko-Kamieńskiej. Od 2018 do 2020 r. asystent, a od 2020 r. adiunkt w Katedrze Etyki Instytutu Filozofii na Wydziale Filozofii KUL.

Główne obszary badań: problem przedłużania życia ludzkiego, transhumanizm, koncepcja osoby w filozofii pielęgniarstwa

Autor książki *Przedłużanie życia jako problem moralny* (2017).

Jakub G o m u ł k a, doktor habilitowany, profesor AGH, filozof. Studia z zakresu filozofii w Papieskiej Akademii Teologicznej w Krakowie (obecnie Uniwersytet Papieski Jana Pawła II), z zakresu informatyki na Politechnice Opolskiej.

W latach 2009-2017 pracownik Wydziału Filozoficznego UPJPiI, 2017-2022 Instytutu Filozofii i Socjologii Uniwersytetu Pedagogicznego im. KEN w Krakowie. Od 2022 r. profesor uczelni w Katedrze Technologii Informatycznych i Mediów na Wydziale Humanistycznym Akademii Górniczo-Hutniczej im. St. Staszica.

Główne obszary zainteresowań badawczych: filozofia religii, kultury, umysłu, techniki, humanistyka cyfrowa, problematyka sztucznej inteligencji, technologia grafów wiedzy, myśl Ludwiga Wittgensteina i Stanisława Lema.

Najważniejsze publikacje książkowe: *Gramatyka wiary. Dziedzictwo Wittgensteinowskiego fideizmu* (2011); *Rachunek. Filozofia nauk formalnych i jej związek z koncepcją podmiotu we wczesnym i średnim okresie twórczości Ludwiga Wittgensteina* (2016); *Filozoficzny Lem* (t. 1, współred., 2021).

Małgorzata Gruchola, doktor habilitowany, profesor KUL, socjolog, medioznawca, kulturoznawca. Studia na Uniwersytecie Marii Curie-Skłodowskiej i na Katolickim Uniwersytecie Lubelskim (doktoranckie – socjologia).

Od 2008 r. pracownik KUL, obecnie profesor uczelni w Katedrze Komunikacji Wizualnej i Nowych Mediów Instytutu Dziennikarstwa i Zarządzania na Wydziale Nauk Społecznych. Ponadto od 2023 r. profesor Instytutu Badań Edukacyjnych (IBE) w Warszawie.

Członek Association for Slavic, East European, and Eurasian Studies (ASEEES) w Chicago, Polskiego Towarzystwa Badań nad Filmem i Mediami, Polskiego Towarzystwa Edukacji Medialnej, Polskiego Towarzystwa Komunikacji Społecznej, Polskiego Towarzystwa Kulturoznawczego. Członek korespondent Towarzystwa Naukowego KUL.

Główne obszary zainteresowań badawczych: społeczno-kulturowe determinanty oddziaływania kolejnych etapów nowych technologii informacyjno-komunikacyjnych (Web 1.0 – Web 5.0, internet rzeczy, internet semantyczny) oraz technologii sztucznej inteligencji, aksjologia mediów i AI, projekty technologicznego ulepszania człowieka w świetle personalizmu chrześcijańskiego, edukacja medialna, socjologia kultury medialnej.

Najważniejsze publikacje książkowe: *Ochrona telewizja w III Rzeczypospolitej* (2003); *Gmina Kock dawniej i dziś* (2006); *Polityka społeczna w gminie Kock* (2008); *Ochrona użytkowników internetu w państwach Unii Europejskiej* (2012); *Władysław Froch. W poszukiwaniu tożsamości...* (2017); *Kompetencje medialne w komunikacji wizualnej* (współautor, 2020).

Leszek Korporowicz, profesor, socjolog. Studia z zakresu socjologii na Uniwersytecie Warszawskim.

Od 1983 do 2008 r. pracownik Uniwersytetu Warszawskiego, w latach 1992-1994 wicedyrektor Instytutu Socjologii, od 1997 profesor. W latach 2000-2008 kolejno prorektor i rektor Szkoły Wyższej Bogdana Jańskiego. Od 2008 r. zatrudniony na stanowisku profesora na Uniwersytecie Jagiellońskim, m.in. jako dyrektor Instytutu Bliskiego i Dalekiego Wschodu (2010-2012), kierownik Zakładu Komunikacji Międzykulturowej (2009-2012), Katedry Antropologii Kultury Współczesnej (2019-2021) i Zakładu Humanistycznych Problemów Nauk o Kulturze (2021-2023) na Wydziale Studiów Międzynarodowych i Politycznych UJ.

Współzałożyciel Polskiego Towarzystwa Ewaluacyjnego (w latach 2000-2002 prezes Zarządu Głównego, obecnie prezes honorowy), członek Zespołu Pedagogiki Kultury i Edukacji Międzynarodowej Komitetu Nauk Pedagogicznych PAN, Stowarzyszenia Wspierania Edukacji Międzykulturowej, Polskiego Towarzystwa Socjologicznego, Polskiego Towarzystwa Kulturoznawczego.

Główne obszary zainteresowań badawczych: socjologia kulturowa, komunikacja i relacje międzykulturowe, antropologia kultury współczesnej, dziedzictwo kulturowe, ewaluacja.

Najważniejsze publikacje książkowe: *Tworzenie sensu* (1996); *Komunikacja międzykulturowa. Zderzenia i spotkania* (współred., 1996); *Osobowość i komunikacja w społeczeństwie transformacji* (1997); *Socjologia kulturowa. Kontynuacje i poszukiwania* (2011); *Mosty nadziei. Jagiellońskie inspiracje dialogu międzykulturowego* (2016). *Jagiellonian Ideas Towards Modern Times* (współred. 2017); *Studia kulturowe* (współred. 2022), animator i redaktor trzech tomów serii *Jagiellonian Cultural Studies* (2012, t. 1-2; 2016).

Roman M. K r z a n o w s k i, doktor, filozof, informatyk. Studia z zakresu sztucznej inteligencji w Birkbeck College i na University of London, z zakresu filozofii na Uniwersytecie Papieskim Jana Pawła II w Krakowie.

Od 2020 r. adiunkt w Katedrze Filozofii na UPJP2.

Członek m.in. Association for the Advancement of Artificial Intelligence (AAAI) i Komisji Filozofii Nauk PAU.

Główne obszary zainteresowań badawczych: filozofia informatyki i informacji, filozoficzne podstawy sztucznej inteligencji, etyka i problemy etyczne w społeczeństwie informacyjnym, etyka robotyczna, ontologia, filozofia w technologii, technologie sieciowe, systemy informacji przestrzennej, przetwarzanie informacji.

Najważniejsze publikacje książkowe: *Metro Ethernet Services for LTE* (2014); *Studying with Grand Master: The Art of Tae Kwon Do* (2014); *Spatial Evolutionary Modeling* (współautor, 2001); *Ontological Information* (2022). Współautor międzynarodowych standardów sieciowych.

Anna M a j, doktor habilitowany, kulturoznawca, medioznawca, komunikolog. Studia z zakresu kulturoznawstwa na Uniwersytecie Śląskim.

Od 2006 r. pracownik UŚ, obecnie adiunkt w Instytucie Nauk o Kulturze na Wydziale Humanistycznym.

Członek International American Studies Association (IASA), International Communication Research & Education Association (ECREA), Media Art Histories Association (MAH), Polskiego Towarzystwa Edukacji Medialnej (PTM).

Główne obszary zainteresowań badawczych: nowe media, antropologia mediów, Science, Technology and Society (STS), teoria komunikacji, sztuka i komunikacja cyfrowa, teoria wiedzy, teoria podróży, współczesne praktyki kulturowe.

Najważniejsze publikacje książkowe: *Media w podróży* (2008); *Digital Memories: Exploring Critical Issues* (współred., 2009); *Kody McLuhana. Topografia nowych mediów* (współred., 2009); *The Real and the Virtual: Critical Issues in Cybercultures* (współred., 2009); *Emerging Practices in Cyberculture and Social Networking* (współred., 2010); *Post-Privacy Culture: Gaining Social Power in Cyber-Democracy* (red., 2014); *Cyberculture Now. Social and Communication Behaviours on the Web* (red., 2019); *Przemiany wiedzy w cyberkulturze. Badania nad kulturą, komunikacją, wiedzą i mediami* (2021).

Zbigniew O r b i k, doktor habilitowany, profesor PŚ, filozof. Studia z zakresu filozofii na Uniwersytecie Jagiellońskim.

Od 2002 r. zatrudniony na Politechnice Śląskiej, obecnie profesor uczelni w Katedrze Stosowanych Nauk Społecznych na Wydziale Organizacji i Zarządzania. Ponadto pracownik Wydział Psychologii i Nauk o Rodzinie w Krakowskiej Akademii im. A.F. Modrzewskiego (2002-2010), Szkoły Wyższej im. Bogdana Jańskiego w Opolu (2010-2016) oraz Wydziału Zarządzania i Komunikacji Społecznej Uniwersytetu Jagiellońskiego (2016-2017).

Członek Oddziału Krakowskiego Polskiego Towarzystwa Filozoficznego.

Główne obszary zainteresowań badawczych: historia szkoły lwowsko-warszawskiej, filozofia nauki, filozofia świadomości, etyka.

Najważniejsze publikacje książkowe: *Czas i świadomość w filozofii Bergsona i Husserla* (2017); *Filozofia Izydory Dąmbskiej* (2018).

Grażyna O s i k a, doktor habilitowany, profesor PŚ, filozof. Studia z zakresu filozofii na Uniwersytecie Śląskim.

Od 1996 zatrudniona na PŚ, obecnie profesor uczelni w Katedrze Stosowanych Nauk Społecznych na Wydziale Organizacji i Zarządzania. Współpracownik Gliwickiej Wyższej Szkoły Przed-

siębiorczości, Górnośląskiej Wyższej Szkoły Handlowej im. Józefa Korfantego w Katowicach oraz Uniwersytetu Śląskiego (od 2020 r. na Wydziale Sztuki i Nauk o Edukacji w Cieszynie).

Członek Polskiego Towarzystwa Komunikacji Społecznej (od 2019 r. przewodnicząca sekcji Filozofia i Antropologia Komunikacji).

Główne obszary zainteresowań badawczych: pogranicza filozofii i komunikacji społecznej, w tym filozofia mediów, filozofia komunikacji, ekologia hipertekstu, tożsamość osobowa, kompetencje komunikacyjne, wpływ technologii cyfrowej na społeczeństwo.

Najważniejsze publikacje książkowe: *Bunt i reforma* (red., 2011); *Procesy i akty komunikacyjne. Koncepcje klasyczne i współczesne* (2011); *Tożsamość osobowa w epoce cyfrowych technologii komunikacyjnych* (2016).

Paweł P o l a k, doktor habilitowany, profesor UPJPII, filozof. Studia z zakresu elektroniki i telekomunikacji w Akademii Górniczo-Hutniczej w Krakowie, Wydział Elektrotechniki, Automatyki, Informatyki i Elektroniki, z zakresu filozofii (doktoranckie) w Papieskiej Akademii Teologicznej w Krakowie (obecnie Uniwersytet Papieski Jana Pawła II).

Od 2004 r. pracownik PAT, obecnie profesor uczelni w Katedrze Historii i Filozofii Nauki na Wydziale Filozoficznym UPJPII.

Członek Centrum Kopernika Badań Interdyscyplinarnych (Kraków), Society for Philosophy and Technology, Komisji Filozofii Nauki Polskiej Akademii Umiejętności (w latach 2015-2021 sekretarz. 2021-2024 wiceprzewodniczący). W latach 2020-2023 członek-ekspert Komitetu Historii Nauki i Techniki PAN.

Główne obszary zainteresowań badawczych: filozofia techniki, filozofia sztucznej inteligencji, historia i filozofia nauki, historia filozofii polskiej.

Najważniejsze publikacje książkowe: (rok pierwszego wydania), a w przypadku ich braku – ogólnie – tematyka publikowanych artykułów.

Najważniejsze książki: *Dynamika nauki: filozoficzne aspekty modelowania rozwoju nauki przy pomocy układów dynamicznych* (2004); „Byłem Pana przeciwnikiem [profesorze Einstein]...”. *Relatywistyczna rewolucja naukowa z perspektywy środowiska naukowo-filozoficznego przedwojennego Lwowa* (2012); *Metodyka pracy naukowej w filozofii. Wprowadzenie do warsztatu cyfrowej humanistyki* (współautor, 2023).

Łukasz S a r o w s k i, doktor, medioznawca, filozof. Studia z zakresu kulturoznawstwa i filozofii (doktoranckie) na Katolickim Uniwersytecie Lubelskim Jana Pawła II.

Od 2021 r. asystent w Katedrze Języka, Retoryki i Prawa Mediów Instytutu Dziennikarstwa i Zarządzania na Wydziale Nauk Społecznych KUL.

Główne obszary zainteresowań badawczych: badania z zakresu ewolucji internetu w kontekście interakcji społecznych, interakcji człowiek-robot oraz wpływu technologii sztucznej inteligencji na zmiany w obrębie komunikacji społecznej.

Autor artykułów z powyższych dziedzin.

Renata T o m a s z e w s k a, doktor habilitowana, profesor UKW, pedagog. Studia z zakresu pedagogiki w Akademii Bydgoskiej im. Kazimierza Wielkiego (obecnie Uniwersytet Kazimierza Wielkiego).

Od 2005 r. pracownik UKW, obecnie profesor uczelni w Katedrze Pedagogiki Pracy i Andragogiki na Wydziale Pedagogiki.

Zastępczyni redaktora naczelnego czasopisma „Szkoła – Zawód – Praca”.

Sekretarz Sekcji Pedagogiki Pracy przy Komitecie Nauk Pedagogicznych PAN, koordynatorka wchodzącego w skład UNESCO International Centre for Technical and Vocational Education and

Training (UNEVOC Centre), członkini *International Society for Quality-of-Life Studies* (ISQOLS) i *European Society for Research on the Education of Adults* (ESREA).

Główne obszary zainteresowań badawczych: problematyka czwartej rewolucji przemysłowej i społeczeństwa 5.0, pedagogiki sztucznej inteligencji, edukacji i pracy w XXI wieku, podmiotowości i sprawczości, zjawisk dysfunkcyjnych w obszarze pracy (w szczególności niewolnictwa i handlu ludźmi), relacji zachodzących między pracą zawodową a życiem pozazawodowym.

Najważniejsze publikacje książkowe: *Edukacja w zakładzie pracy w perspektywie organizacji uczącej się* (2012); *Professional education in the context of knowledge based economy* (red., 2013); *Relacje praca–życie pozazawodowe drogą do zrównoważonego rozwoju jednostki* (red., 2014); *Work – personal life. Between harmony and disintegration* (red., 2015); *Edukacja dla przyszłości w perspektywie orientacji na rynek pracy. Debaty, szkice, refleksje polsko-ukraińskie* (współred., 2017); *Wokół podstawowych zagadnień pedagogiki pracy* (współred., 2017); *Praca zawodowa – życie osobiste. Dysonans czy synergia?* (2018); *Generacja NEET. Obraz własny. Doświadczenia pracodawców i instytucji rynku pracy* (współautorka, 2019); *Sekrety organizacji. Barwy codzienności* (red., 2019); *Praca i kariera w warunkach nowego millenium. Implikacje dla edukacji* (red., 2020); *Człowiek i praca. Perspektywa transhumanizmu* (e-book, t. 1, 2021); *Człowiek w organizacji* (współautorka, t. 1, 2022, t. 2, 2023).

Mariusz W o j e w o d a, doktor habilitowany, profesor UŚ, filozof. Studia z zakresu filozofii na Uniwersytecie Śląskim i Uniwersytecie Jagiellońskim (doktorat), z zakresu teologii na Uniwersytecie Opolskim.

Od roku 1992 zatrudniony na UŚ, obecnie profesor uczelni w Instytucie Filozofii na Wydziale Humanistycznym, kierownik zespołu badawczego „Filozofia wobec cywilizacyjnych wyzwań współczesności” (FwCWW). Ponadto od 2006 r. pracownik Uniwersytetu WSB Merito w Chorzowie, a od 2016 Akademii Humanitas w Sosnowcu.

Członek Słowackiego Towarzystwa Filozoficznego, Polskiego Towarzystwa Filozoficznego, Polskiego Towarzystwa Komunikacji Społecznej. Członek Międzynarodowej Grupy Badawczej Cultmedia na Uniwersytecie w Poczdamie oraz Komisji Etyki Badań Naukowych na UŚ.

Główne zainteresowania badawcze: aksjologia, etyka, etyki stosowane, filozofia mediów, filozofia techniki.

Najważniejsze publikacje książkowe: *Wartości – tradycja i współczesność* (współred., 2009); *Pluralizm aksjologiczny i jego implikacje we współczesnej filozofii religii* (2010); *Filozofia w służbie polityki* (współred., 2011); *Problem ładu społecznego w społeczeństwie wielokulturowym* (współred., 2011); *Podstawy filozofii dla uczniów i studentów* (cz. 3, współautor, 2012); *Być, mieć, czy władać* (współred., 2013); *Die Zukunft von Medien-Räumen. Zwischen Freiheit und Zwängen* (współred., 2017).

CONTENTS
Ethos 36, no. 3 (143) (2023)

DOMESTICATING ARTIFICIAL INTELLIGENCE

From the Editors, Lęk i nadzieja (A.L.K., M.W.)	5
From the Editors, Fear and Hope (A.L.K., M.W.)	13
J a n P a w e ł II, Szybki rozwój (Fragmenty Listu apostolskiego do odpowiedzialnych za środki społecznego przekazu, Watykan, 24 I 2005)	23
J o h n P a u l II, The Rapid Development (Extracts from the Apostolic Letter to Those Responsible for Communications, Vatican, 24 January 2005)	33

TOWARDS AN UNDERSTANDING OF ARTIFICIAL INTELLIGENCE

Zbigniew O r b i k, In Search of Philosophical Sources of the Problem of Artificial Intelligence	45
DOI 10.12887/36-2023-3-143-05	
Fr. Marcin F e r d y n u s, Do We Have the Right to “Tame” Artificial Intelligence?	67
DOI 10.12887/36-2023-3-143-06	
Paweł P o l a k, Ryszard K r z a n o w s k i, How to Tame Artificial Intelligence? A Symbiotic Model for Beneficial AI	92
DOI 10.12887/36-2023-3-143-07	

ARTIFICIAL INTELLIGENCE IN HUMAN SOCIETY

Mariusz W o j e w o d a, Artificial Intelligence as a Social Utopia	109
DOI 10.12887/36-2023-3-143-08	
Grażyna O s i k a, The Communication of Human Beings with Artificial Intelligence: Boundary Conditions	127
DOI 10.12887/36-2023-3-143-09	

Łukasz S a r o w s k i, Małgorzata G r u c h o ł a, Social Robotics in the Perspective of (Non) Anthropocentrism.	154
DOI 10.12887/36-2023-3-143-10	
Anna M a j, The Evolution of Intelligence and Its Research: Taming Artificial Intelligence, as seen from the Perspective of Changes in Culture and Communication	177
DOI 10.12887/36-2023-3-143-11	

VISIONS OF THE FUTURE

Jakub G o m u ł k a, In Search of Philosophical Sources of the Problem of Artificial Intelligence . . .	207
DOI 10.12887/36-2023-3-143-12	
Leszek K o r p o r o w i c z, Reductive Pitfalls of the Sociology of “Thinking Machines”	235
DOI 10.12887/36-2023-3-143-13	
Renata T o m a s z e w s k a, Paradise—Hell—Triumph: Artificial Intelligence in Scenarios for the Future	253
DOI 10.12887/36-2023-3-143-14	

THINKING ABOUT THE FATHERLAND...

Marek B e r n a c k i, “Time future will embrace us with one name / That of the imperfect and the already forgotten.” On the Unknown Juvenile Poems by Czesław Miłosz	277
DOI 10.12887/36-2023-3-143-15	

NOTES AND REVIEWS

Robert K u b l i k o w s k i, A Philosophy of Artificial Intelligence (review of M. A. Boden, <i>Sztuczna inteligencja: Jej natura i przyszłość; Krótkie wprowadzenie</i> , trans. T. Sieczkowski, Łódź: Wydawnictwo Uniwersytetu Łódzkiego, 2020).	293
Books recommended by <i>Ethos</i> (J. Landgrebe, B. Smith, <i>Why Machines Will Never Rule the World: Artificial Intelligence without Fear</i> , New York – London: Routledge, 2022)	298

THROUGH THE PRISM OF THE ETHOS

Mirosława C h u d a, Anticipation	303
Notes about the Authors.	309

ETHOS
Kwartalnik Instytutu Jana Pawła II KUL
ISSN 0860-8024
Prefix DOI 10.12887

Instytut Jana Pawła II
Wydział Filozofii
Katolicki Uniwersytet Lubelski Jana Pawła II
Al. Raławickie 14, 20-950 Lublin
tel. 81 4453218
fax 81 4453217
e-mail: ethos@kul.lublin.pl

INFORMACJE DLA AUTORÓW

Tematy kolejnych numerów monograficznych „Ethosu” są zapowiadane w wyprzedzeniem na czwartej stronie okładki oraz na stronie internetowej kwartalnika. Artykuły związane z tematami poszczególnych numerów można nadsyłać od chwili ukazania się zapowiedzi. Wszystkie artykuły naukowe publikowane w „Ethosie” poddawane są recenzji w trybie anonimowym dla obu stron, dokonywanej przez dwóch niezależnych recenzentów zewnętrznych. Jeśli jedna z recenzji okaże się negatywna, redakcja może podjąć decyzję o skierowaniu tekstu do trzeciego recenzenta. W przypadku, gdy trzecia recenzja jest pozytywna, autor zobowiązany jest do odniesienia się w zmienionej wersji artykułu do uwag wszystkich recenzentów. Jeśli obydwie recenzje są negatywne, artykuł nie zostanie opublikowany.

Redakcja zastrzega sobie prawo do decyzji o nieskierowaniu artykułu do recenzji, a także o rezygnacji z publikacji artykułu na każdym etapie prac redakcyjnych. W sprawach spornych ostateczną decyzję o publikacji artykułu podejmuje redaktor naczelny kwartalnika „Ethos”.

W celu ułatwienia procesu recenzji anonimowej, prosimy nie umieszczać w tekście informacji, które pozwoliłyby zidentyfikować autora.

Przed publikacją wszystkie teksty poddawane są adiustacji językowej i opracowaniu redakcyjnemu zgodnie ze stylem edytorskim czasopisma, a następnie autoryzowane.

ZASADY OGÓLNE

Objętość artykułu, wraz z bibliografią i abstraktami, nie powinna przekraczać 20 stron maszynopisu (1 arkusza wydawniczego – 40 tysięcy znaków ze spacjami), objętość recenzji i sprawozdań – 10 stron maszynopisu (0,5 arkusza wydawniczego – 20 tysięcy znaków ze spacjami).

Do artykułu należy dołączyć abstrakt (o objętości nie większej niż 1 strona maszynopisu – 1800 znaków ze spacjami) i słowa kluczowe w języku polskim i angielskim, a także bibliografię zawiera-

jąca pełne opisy cytowanych pozycji (z rozpisanymi imionami autorów, tłumaczy i redaktorów oraz pełnymi nazwami wydawnictw).

Abstrakt powinien być napisany w metajęzyku i zawierać informacje, jaki jest cel artykułu i jaki problem został w nim poruszony, jakie metody badawcze autor wykorzystał, do jakich wniosków doszedł i w jakich kierunkach badania przedstawione w tekście mogą być kontynuowane w przyszłości.

W przypadku tekstów nadsyłanych w języku angielskim prosimy stosować zasady edytorskie rekomendowane przez The Chicago Manual of Style.

Artykuły niedostosowane do stylu edytorskiego „Ethosu” nie będą przyjmowane.

Autorzy zobowiązani są do wypełnienia i przesłania (wraz z artykułem) do redakcji (1) oświadczenia o oryginalności dzieła oraz (2) umowy o dzieło i przeniesienie autorskich praw majątkowych, a także (3) formularza, na podstawie którego opracowywane są ich noty biograficzne.

Złożenie artykułu do publikacji jest rozumiane jako deklaracja, że tekst ten nie został nigdzie wcześniej opublikowany również w postaci elektronicznej oraz że nie jest on równocześnie złożony w redakcjach innych czasopism ani rozpatrywany jako fragment monografii.

Zasadniczo nie zamieszczamy materiałów, które ukazały się wcześniej w druku lub w postaci publikacji elektronicznej. Wyjątkiem są tłumaczenia z języków obcych tekstów dotąd niedostępnych w języku polskim.

Tekst wykładu lub odczytu powinien zawierać informację o tym, kiedy i na jakim forum był prezentowany.

Autorzy proszeni są o ujawnianie wkładu wszystkich osób, które przyczyniły się powstania składanego do publikacji artykułu (z podaniem ich afiliacji oraz sprecyzowaniem ich udziału), oraz informacji o źródłach finansowania badań przeprowadzonych w związku z jego przygotowywaniem.

Ghostwriting i guest authorship są przejawem nierzetelności naukowej, a wszelkie wykryte przypadki będą demaskowane, włącznie z powiadomieniem odpowiednich podmiotów. Wszelkie przejawy nierzetelności naukowej będą dokumentowane.

FORMAT PLIKU

Prosimy o składanie artykułów w postaci elektronicznej, w formacie doc lub docx. W przypadku, gdy praca na pliku przygotowanym za pomocą innego edytora niż Microsoft Word okazuje się utrudniona ze względów technicznych, redakcja zastrzega sobie prawo do rezygnacji z publikacji tekstu.

FORMAT DOKUMENTU

Formatowanie tekstu prosimy ograniczyć do minimum. Prosimy na żadnym etapie pracy na pliku nie stosować specjalnych „stylów” dostępnych w edytorze tekstu, nie zamieszczać komentarzy „w dymkach” ani nie posługiwać się opcjami: „porównaj dokumenty”, „pokaż adiustację” ani „śledź zmiany”.

W całym dokumencie prosimy stosować czcionkę Times New Roman 12 pkt i nie wydzielać cytatów za pomocą innego kroju, rozmiaru czcionki ani zmiany szerokości kolumny.

W artykułach należy stosować alfabet łaciński, a w przypadku słów pochodzących z języków posługujących się innym alfabetem używać transliteracji/transkrypcji.

Kursywę prosimy stosować wyłącznie w tytułach cytowanych publikacji. Słów obcojęzycznych prosimy nie zapisywać kursywą.

Preferowana interlinia: 1,5.

Wcięcia akapitowe prosimy wprowadzać za pomocą tabulatora (a nie spacji).

W całym tekście prosimy wprowadzić numerację stron.

Przypisy należy umieszczać na dole strony (przypisów końcowych nie stosujemy). Przypisy powinny zawierać informacje bibliograficzne; wszelkie informacje oraz argumentacja powinny znajdować się w tekście głównym.

FORMAT PRZYPISÓW

Gwoli jednolitości edytorskiej, którą chcemy utrzymać, mimo że pismo ma charakter interdyscyplinarny, a w różnych dziedzinach stosowane są różne zapisy bibliograficzne, prosimy wszystkich autorów o stosowanie następujących reguł:

Przypisy, w których po raz pierwszy przywoływane jest określone dzieło, powinny zawierać pełne dane bibliograficzne, według wzoru:

pozycja książkowa (wymagane dane: autor, tytuł, tłumacz, redaktor, wydawnictwo, miejsce i rok wydania),

np. R. Buttiglione, *Mysł Karola Wojtyły*, tłum. J. Merecki SDS, Instytut Jana Pawła II KUL, Lublin 2010.

artykuł w czasopiśmie (wymagane dane: nazwisko autora, tytuł, tytuł czasopisma, tom/rocznik, (w nawiasie) rok, a następnie numer bieżący i (w nawiasie) ciągły oraz numery stron),

np. G.E.M. Anscombe, *Nowożytna filozofia moralności*, tłum. M. Roszyk, „Ethos” 23(2010) nr 4(92), s. 39-60.

artykuł w publikacji zbiorowej (wymagane dane: autor, tytuł, tłumacz, tytuł publikacji zbiorowej, tłumacz(e), redaktor(zy), wydawnictwo, miejsce i rok wydania oraz numery stron), np. S. Judycki, *Istnienie i natura duszy ludzkiej*, w: *Antropologia*, red. S. Janeczek, Wydawnictwo KUL, Lublin 2010, s. 121-177.

W przypadku cytatu prosimy o podawanie wyłącznie numeru strony, z której cytat pochodzi. W przypadku odwoływania się do fragmentu tekstu prosimy zamieszczać informację, na których stronach fragment ten się znajduje.

Kiedy dzieło cytowane jest po raz kolejny, prosimy stosować zapis skrócony: nazwisko autora, dz. cyt., numer strony,
np. Buttiglione, dz. cyt., s. 10.

Jeśli wcześniej cytowane są również inne dzieła danego autora, po nazwisku autora prosimy podać tytuł dzieła,
np. Buttiglione, *Myśl Karola Wojtyły*, s. 10.

W przypadku dzieł klasycznych, w których stosowana jest wewnętrzna notacja, prosimy zamieścić ją po tytule dzieła,
np. Platon, *Timajos*, 86 B-C, w: tenże, *Timajos, Kritias*, tłum. W. Witwicki, Antyk, Kęty 2002, s. 85.

Cytaty z nauczania Jana Pawła II prosimy przytaczać za polskim wydaniem „L'Osservatore Romano”, *Nauczaniem papieskim* (red. E. Weron SAC, A. Jaroch SAC, Pallottinum) lub za wydaniem watykańskimi.

Skrót zob. odnosi czytelnika do całości dzieła, a skrót por. do jego fragmentu (wymagamy wówczas podania zakresu numerów stron).

W przypadku cytowania haseł ze słowników bądź wydawnictw o charakterze encyklopedycznym prosimy o następujący zapis: autor, hasło „Tytuł hasła”, w: tytuł publikacji, redaktor, wydawnictwo, miejsce i rok wydania, numery stron (kolumny, szpalty etc.).

W przypadku klasycznych zbiorów tekstów źródłowych, w których nie ma numeracji stron i we wszystkich wydaniach zachowana jest ta sama numeracja, np. kolumn, dopuszczamy podanie wyłącznie pełnego tytułu i numeru, np. kolumny.

Prosimy nie stosować skrótów: passim, j.w. ani znaku =.

Prosimy rozpisywać wszystkie skróty, zastępować cyfry rzymskie łacińskimi i nie podawać numerów wydań cytowanych pozycji.

CYTOWANIE Z TEKSTÓW OBCOJĘZYCZNYCH

Cytaty z tekstów obcojęzycznych powinny pochodzić z dostępnych przekładów na język polski. W przypadku braku źródła polskojęzycznego należy zaznaczyć, że tłumaczem tekstu jest autor artykułu bądź podać nazwisko tłumacza. Polskie przekłady (polskie wersje językowe) uważamy za podstawę odniesień i przestrzegamy zasady, że oryginalne wersje językowe przywoływane są tylko wtedy, gdy między nimi a polskimi przekładami zachodzą istotne różnice. W przypadkach takich dopuszczalne jest wprowadzenie tłumaczenia własnego przy zachowaniu informacji bibliograficznej na temat dostępnego polskiego przekładu.

W przypadku braku polskiego tłumaczenia danego dzieła prosimy cytować jego fragmenty za tekstem oryginalnym, a nie za jego obcojęzycznymi przekładami, np. nie należy cytować tekstów Hume'a za ich przekładami na język włoski.

ILUSTRACJE

Autorzy zobowiązani są uzyskać zgodę na publikację towarzyszących artykułom materiałów objętych prawem autorskim (np. fotografii, faksymile, reprodukcji dzieł sztuki). Prosimy dołączać pliki graficzne o rozdzielczości 300 dpi (preferowany format TIF). Każda ilustracja powinna zostać opatrzona podpisem zawierającym informacje o jej treści, źródle oraz wszelkie inne informacje wymagane prawem autorskim.

PRAWA AUTORSKIE

Złożenie tekstu do druku jest równoznaczne z przekazaniem wyłącznych praw autorskich kwartalnikowi „Ethos”. W przypadku zamiaru przedruku tekstu opublikowanego wcześniej w „Ethosie” autor (lub w jego imieniu wydawca) proszony jest o zwrócenie się o zgodę do Instytutu Jana Pawła II KUL. Prośby takie są z reguły rozpatrywane pozytywnie z zastrzeżeniem, że ponownej publikacji towarzyszyć będzie informacja bibliograficzna na temat pierwodruku.

Prawa autorskie na przedruk, przekład bądź wykorzystanie w formie elektronicznej materiału pierwotnie opublikowanego w „Ethosie”, jeśli występuje o nie podmiot inny niż autor materiału, udzielane są odpłatnie. Składając prośbę o udzielenie praw autorskich, prosimy określić charakter publikacji, jej język oraz nakład. „Ethos” nie we wszystkich przypadkach jest właścicielem praw autorskich do opublikowanych w piśmie materiałów. Przed złożeniem wniosku prosimy zwrócić uwagę na informacje zawarte w przypisach.

ETHOS
QUARTERLY OF THE JOHN PAUL II INSTITUTE
ISSN 0860-8024
DOI prefix: 10.12887

The John Paul II Institute
Faculty of Philosophy
The John Paul II Catholic University of Lublin
Al. Raławickie 14, 20-950 Lublin, Poland
Phone: +48 81 4453218
Fax +48 81 4453217
E-mail: ethos@kul.lublin.pl

SUBMISSION GUIDELINES

The themes of the forthcoming volumes of *Ethos* can be found on the back cover of every volume and on the journal's website. Submissions are subject to double-blind peer review by two independent and external referees and to editorial revisions. One of the referees will represent an academic milieu from outside Poland, non-related to that of the author; the other will be a Polish scholar. In order to facilitate double-blind peer review process, authors are requested to eliminate all references that might identify them.

The editors reserve the right to decide whether a received manuscript will be considered for publication.

GENERAL GUIDELINES

The suggested length for *Ethos* articles is 4,000 to 8,000 words, including notes. Review and report articles should not exceed 4,000 words.

An article should be followed by a bibliography comprising a full list of the references, an abstract (up to 1,800 characters), and keywords.

The abstract should define the goal of the paper and outline the problem it discusses, as well as the methods applied throughout the scrutiny, the conclusions the author has reached, and the proposed further research on the subject.

Authors are requested to fill in a form to provide information necessary for the editors to prepare biographical notes about the contributors.

Submission of an article written in a language other than Polish should be accompanied by a declaration that the text in question has not been published elsewhere, either in print or online, and that it is not being considered for publication in any other periodical or monograph.

Authors are requested to disclose names of anyone participating in the preparation of the submission (together with their affiliation and contribution), as well as all the sources of research funding used in its preparation.

Ghostwriting and guest authorship are violation of the standard codes of scholarly conduct, and all its instances will be unmasked and documented, and the appropriate entities notified.

FORMATTING THE FILE

Please submit your file as a Word document (doc or docx).

FORMATTING THE DOCUMENT

Lines should be double-spaced, including quotations and notes, using Times New Roman font in 12 point size. Format your paper following the latest edition of the *Chicago Manual of Style*. Place page numbers on each page. Notes should be numbered consecutively and placed at the foot of each page.

Notes should be used primarily for citation purposes and include all the necessary bibliographic information. Notes should not contain any discussion that can be incorporated into the body of the paper.

IMAGES

Authors are expected to obtain permission to reproduce any copyrighted materials (e.g., photographs, facsimiles, art reproductions) used in their articles. Please submit 300 dpi TIF files. Supply each illustration with a caption, accompanied by a source line and any required acknowledgments.

QUOTATIONS FROM FOREIGN LANGUAGES

In cases where standard scholarly translations of quoted texts are available, scholars may rely upon them for both text and notes. If such editions are not readily available, the author must provide the original texts for the use of reviewers and translators.

Please use the Latin alphabet throughout your manuscript and apply transliteration/transcription in cases of words from languages using other alphabets.

CITATIONS FORMAT

The first reference to a given work should include full bibliographic information, as in the following basic citation examples:

(book)

Rocco Buttiglione, *Mysł Karola Wojtyły*, trans. Jarosław Merecki (Lublin: Instytut Jana Pawła II KUL, 2010).

(journal article)

Edward Sapir, "Sound Patterns in Language," *Language* 1, no. 2 (1925): 37–51.

(article in a monograph)

Stanisław Judycki, "Istnienie i natura duszy ludzkiej," in: *Antropologia*, ed. Stanisław Janeczek (Lublin: Wydawnictwo KUL, 2010), 121–77.

In cases of direct quotations, please supply only the number of the page from which the given quotation comes.

When a given work has already been cited, please shorten the reference to: author, title, page number.

References to the classical texts must include the standard identification numbers for those texts, e.g., Plato, *Timaeus*, 86 B-C.

References to the speeches of John Paul II should include precise information on when and where a given speech was delivered. References to encyclicals and other Catholic Church documents should include the numbers of the sections referred to.

COPYRIGHT AND PERMISSIONS

By submitting a text for publication, authors grant an exclusive license to the quarterly *Ethos*. Should authors wish to republish their article in another periodical or in a monograph, they are requested to apply for permission to the John Paul II Institute at the Catholic University of Lublin. Such permissions are generally granted on condition that all the required credits are included in the republished version.

If you wish to reproduce any material published in *Ethos* in a paper or in an electronic work that is not the author's own, please contact the John Paul II Institute at the Catholic University of Lublin. Permissions and rights are granted after processing your request and involve fee payment. Your request should include a description of the intended publication, its language and print run. Note that *Ethos* may not be the copyright holder of some of the material published in the journal. Please check the copyright or acknowledgment notes before making your request.

Wydawca

WYDZIAŁ FILOZOFII KUL
OŚRODEK BADAŃ NAD MYŚLĄ JANA PAWŁA II
„INSTYTUT JANA PAWŁA II KUL”
Al. Raławickie 14, 20-950 Lublin
telefon: +48 81 4453217, telefon redakcji: +48 81 4453218
e-mail: ethos@kul.lublin.pl

Instytut Jana Pawła II powołany został do istnienia uchwałą Senatu Akademickiego KUL 25 VI 1982 r. jako „międzywydziałowy ośrodek naukowy i dydaktyczny Katolickiego Uniwersytetu Lubelskiego”, przeznaczony do studiów „myśli i dzieła papieża Jana Pawła II” oraz do „budowania wspólnoty osób w duchu głoszonej przez Niego nauki Chrystusa” (Statut Instytutu). Działalność Instytutu koncentruje się na filozofii i teologii człowieka, moralności oraz misji duszpasterskiej św. Jana Pawła II. Instytut realizuje swoje zadania głównie poprzez doroczne sympozja naukowe, działalność badawczą i dydaktyczną (konwersatoria) oraz wydawniczą. Instytut Jana Pawła II stanowią: Rada Naukowa (przewodniczący – prof. dr hab. Andrzej Derdziuk OFMCap), Zarząd Instytutu (dyrektor – ks. dr hab. Marek Słomka, prof. KUL), pracownice oraz kwartalnik „Ethos”.

**Rada Naukowa
Instytutu
Jana Pawła II**

Prof. Dr hab. Andrzej DERDZIUK OFMCap – przewodniczący Rady Naukowej
Ks. Prof. Dr hab. Mirosław KALINOWSKI – rektor KUL
Ks. Prof. Dr hab. Mirosław SITARZ – prorektor KUL
S. Prof. Dr hab. Beata ZARZYCKA – prorektor KUL
Prof. Dr hab. Beata PISKORSKA – prorektor KUL
Prof. Dr hab. Ewa TRZASKOWSKA – prorektor KUL
Prof. Dr hab. Adam ZADROGA – prorektor KUL
Prof. Dr hab. Barbara CHYROWICZ SSpS
Ks. Prof. Dr hab. Tadeusz DOLA
Prof. Dr hab. Alicja GRZEŠKOWIAK
Prof. Dr hab. Wojciech KACZMAREK
Prof. Dr hab. Karol KLAUZA
Prof. Dr hab. Jarosław KUPCZAK OP
Prof. Dr hab. Ewa Agnieszka LEKKA-KOWALIK
Prof. Dr hab. Małgorzata U. MAZURCZAK
Abp Prof. Dr hab. Alfons NOSSOL
Prof. Dr hab. Robert PIŁAT
Prof. Dr hab. Grzegorz PRZEBINDA
Prof. Dr hab. Jacek SALIJ OP
Prof. Dr hab. Stefan SAWICKI
Ks. Prof. Dr hab. Jan SOCHOŃ
Prof. Dr hab. Andrzej STOFF
Ks. Prof. Dr hab. Stanisław SUWIŃSKI
Prof. Dr hab. Władysław STRÓŻEWSKI
Prof. Dr hab. Krzysztof WIAK
Ks. Prof. Dr hab. Alfred M. WIERZBICKI
Ks. Prof. Dr hab. Henryk WITCZYK
Prof. Krzysztof ZANUSSI
Prof. Dr hab. Zofia ZDYBICKA USJK

**Warunki
prenumeraty****Cena prenumeraty krajowej na rok 2023: rocznie 100,00 zł.**

Prenumeratorzy indywidualni oraz instytucje mogą zamawiać prenumeratę, wpłacając wymienioną kwotę na konto 94 1050 1953 1000 0024 3072 5941 z zaznaczeniem nazwy rachunku: „Prenumerata ETHOSU”.

Prenumeratę zagraniczną można zamawiać, wpłacając odpowiednią kwotę na konto: 33 1050 1953 1000 0090 3079 8129 (USD)
95 1050 1953 1000 0090 8066 3355 (EUR)

z zaznaczeniem nazwy rachunku: „Prenumerata ETHOSU”.

Cena pojedynczego numeru – 17 USD (16 EUR), prenumerata roczna – 55 USD (50 EUR).

Kolportaż

Kwartalnik ETHOS (Tomasz Górka), Al. Raławickie 14, 20-950 Lublin
telefon: +48 81 4453217, e-mail: eth-kolp@kul.pl

W poprzednich numerach:

Stanisław Wyspiański: dramat i ethos (48)
Chrześcijaństwo na progu trzeciego tysiąclecia (49-50)
Polskie przełomy (51)
Przesłanie Zbigniewa Herberta (52)
Etyka a demografia (55)
Dzieje: nie-Boska komedia? (56)
O solidarność rodziny narodów Europy (57-58)
O solidarność rodziny narodów świata (59-60)
Prawo niedoskonałe czy niesprawiedliwe? (61-62)
Ethos Pielgrzyma (63-64)
Dramat Judasza (65-66)
Polska Jana Pawła II (67-68)
Osoba w społeczeństwie informacyjnym (69-70)
Wojna sprawiedliwa? (71-72)
Muzyka – piękno, dobro, świętość (73-74)
O nowej edukacji (75)
O filozofii Karola Wojtyły (76)
Ethos teatru (77-78)
Białoruś i Ukraina – bliższa Europa (81)
Ciało – osoba – kultura (82-83)
„I będą oboje jednym ciałem” (84)
Oblicza emigracji (87-88)
Kino i transcendencja (89)
O winie i karze (90-91)
Między cnotą a niegodziwością (92)
O radości (93-94)
Filozof osoby – Tadeusz Styczeń SDS (1931-2010) (95)
Ubóstwo dziś (96)
Ethos słowa (97-98)
Czas i wieczność (99)
Filozofia Vaticanum Secundum (100)
Świadomość i tożsamość (101)
Nasi „bracia mniejsi” (102)
O przyjaźni (103)
Ethos przestrzeni (104)
Świętość – ideał i fakt (105)
O przemocy (106)
Wobec utopii (107)
Lęk (108)
Przyszłość humanistyki (109)
Szaleństwo (110)
Transhumanizm (111)
Ethos filozofii (112)
O ciszy i milczeniu (113)
Ethos wizerunku (114)
Chrześcijaństwo w kulturze Zachodu (115)
Koniec osoby? (116)
Obcy (117)
Wstyd (118)
Ethos światła (119)
Ból (120)
Pieniądz (121)
Krzywda (122)
Książka (123)
Upadek (124)
Słuchanie (125)
Muzyczność (126)
Piękno (127)
Hope (128, tom w języku angielskim)
Rozmowa (129)
Dom (130)
Rzeczy (131)
Władza (132)
Głupota (133)
Język (134)
Pamięć (135)
Ludzie i miejsca (136)
O jedzeniu (137)
Widzenie (138)
Samotność (139)
Ku interpretacji wolności (140)
Wolność – kultura – wartości (141)
Przypadek (142)

W najbliższych numerach:

- * Sztuczna inteligencja w świecie wartości ludzkich
- * Godność
- * Truth and Post-truth (tom w języku angielskim)
- * Harmonie świata
- * Rozum i wiara
- * Oczekiwanie

Cena brutto: 25 zł
w tym VAT: 8%

ISSN 0860-8024



9 770860 802403

www.ethos.lublin.pl