„Ethos" 36(2023) nr 3(143) 253-274
DOI 10.12887/36-2023-3-143-14

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 – Triumf[2], 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. F u r m a n e k, *Człowiek jako obiekt badań humanistycznej pedagogiki pracy*, „Labor et Educatio" 2014, nr 2, s. 16.

[2] Zob. J. G a r r e a u, *Radykalna ewolucja. Czy człowiek udoskonalony przez naukę i technikę będzie jeszcze człowiekiem?*, tłum. A. Kloch, 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 inteligencji[4] 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 inżynieryjnym, 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 punków 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 algorytmizacji[6].

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 superinteligencją[7].

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 efektywny[8]. 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 wynikach[9]. Inteligentne systemy AI rozpoznają wzorce i pamiętają przeszłe wydarzenia, ucząc się z nich. Dzięki temu

[6] Por. A.K. P r z e g a l i ń s k a, *Istoty wirtualne. Jak fenomenologia zmieniała sztuczną inteligencję*, Wydawnictwo Universitas, Kraków 2016, s. 239n.

[7] Zob. hasło „Sztuczna inteligencja", w: *Słownik*, https://www.sztucznainteligencja.org.pl/definicja/sztuczna-inteligencja.

[8] Por. A. P a w l i c k a, M. P a w l i c k i, M. C h o r a ś, *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. A d c o c k, *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 zastosowania[10]. Uczenie maszynowe jest obecnie jedną z najbardziej pożądanych technologii na świecie[11]. 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 praktyce[12]. 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. P a w l i c k a, M. P a w l i c k i, R. T o m a s z e w s k a, M. C h o r a ś, R. G e r l a c h, *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. P a w l i c k a, R. T o m a s z e w s k a, E. K r a u s e, D. J a r o s z e w s k a-C h o r a ś, M. P a w l i c k i, M. C h o r a ś, *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. K r a u s e, R. T o m a s z e w s k a, A. P a w l i c k a, *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 przeznacza 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. P a w l i c k a, P a w l i c k i, C h o r a ś, 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 ludzkości[14].

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 ścisłe 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 implantów, 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. S e j n o w s k i, *Deep learning. Głęboka rewolucja. Kiedy sztuczna inteligencja spotka się z ludzką?*, tłum. P. Cypryański, Wydawnictwo Poltext, Warszawa 2019, s. 15-35.

[14] Por. M. F o r d, *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 rozmaitymi 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 superinteligencji[15].

Prognozom przyszłości związanym z powyższymi przewidywanymi ścieżkami rozwoju towarzyszą oczywiste 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, bojlery, 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-piłkarze 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. Konowrocka-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. W a l s h, *To żyje! Sztuczna inteligencja od logicznego fortepianu po zabójcze roboty*, tłum. W. Sikorski, Wydawnictwo Naukowe PWN, Warszawa 2018, s. 227-243.

[17] Por. J. K a p l a n, *Sztuczna inteligencja. Co każdy powinien wiedzieć*, Wydawnictwo PWN, Warszawa 2019, s. 113-191; J.J. B r y s o n, M.E. D i a m a n t i s, T.D. G r a n t, *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 człowieka[19].

Sztuczna inteligencja rodzi wyzwania także dla doktryn filozoficznych oraz religijnych. Inteligentne 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 pracę[21].

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 technika[22].

---

[19] Por. M. L i p o w i c z, *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 wielekroć 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 ogólnoświatową 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 potomnych[25].

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 Foundation[26]. 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) ziszczą 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 planetoidach[27].

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 nanotechnologię[28], 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 wyrywają 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. D r e x l e r, *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, maleńkie 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 wiedzy[30].

Jak natomiast wskazuje trzeci scenariusz – triumfalny – nie powinniśmy jako ludzie odrzucać zdobyczy nowych technologii, ale musimy mieć głęboką świadomość drzemiącego 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 s c i e n c e - f i c t i o n może okazać się największym błędem w historii[31]. 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. G a r r e a u, dz. cyt., s. 139-193.

[31] Zob. S. H a w k i n g, S. R u s s e l l, M. T e g m a r k, F. W i l c z e k, *Stephen Hawking: Transcendence Looks at the Implications of Artificial Intelligence – But Are We Taking Al 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 rzeczywistości[32]. W związku z tym możliwe będzie też wznoszenie 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 zmagań leży w zdolności do stawania 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 nieprzewidziany kształtować wpływu techniki na ludzką naturę i społeczeństwo[33].

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 narastającym 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 metaversem 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 okażą się zaskakujące”[34].

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. „Załóżmy, 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. Załóżmy, 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 YouTubie 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 Apokalipsy[36].

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 pomocnicy[37].

Z drugiej strony, obecne wektory rewolucji technologicznej mogą prowadzić do fazy przejęcia 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 odnosić[38], 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 zmieni 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 in., *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 misję[39]. 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 twórcami[40].

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 stricte problemem utrzymywania układów sztucznej inteligencji pod kontrolą[41]. 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 wie[42].

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 eseju *Dlaczego przyszłość nas niepotrzebuje?* Bill Joy[43] – 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 Hatalska (por. N. H a t a l 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 skonstruowaniu[44]. 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 ludzie[45].

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 śmiercionośnymi organizmami. Stąd można dywagować, że do sztucznej inteligencji również się przyzwyczaimy, co zresztą już się dzieje[46].

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 stawały 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? Wizje 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, Ferburary 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 punków 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 Łuniewska. 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.

Hatalska, 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 Al 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 trans- i 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. *Deeplearning: Głęboka rewolucja: Kiedy sztuczna inteligencja spotka się z ludzką?* Translated by Piotr Cypryański. Warszawa: Wydawnictwo Poltext, 2019.

Słownik, s.v. “Sztuczna inteligencja.” https://www.sztucznainteligencja.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 punków 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/