

OD REDAKCJI

ROZWIJAĆ CZY NIE ROZWIJAĆ? OTO JEST PYTANIE

Niewiele jest dziś tematów, które tak bardzo rozbudzają i ożywiają zbiorową wyobraźnię, jak zagadnienia związane z rozwojem sztucznej inteligencji (SI, ang. artificial intelligence – AI), często mające posmak sensacji. Ogromne rzesze ludzi śledzą na bieżąco medialne doniesienia o nowych technologiach i nowych obszarach ich zastosowań, a także o kontrowersjach wywoływanych przez nie zawsze w pełni przemyślany i kontrolowany proces wdrażania technologii SI. Powszechnie panuje przekonanie, że w niedalekiej przyszłości dojdzie do radykalnych i zasadniczych zmian, zarówno w najbliższym otoczeniu człowieka, jak też w ludzkim sposobie bycia i działania. Niektóre scenariusze, szkicowane przez takich myślicieli jak na przykład Nick Bostrom¹, Max Tegmark², Ray Kurzweil³, czy Kevin Warwick⁴ zakładają, że w niedługim czasie nastąpi eksplozja superinteligencji, która osiągnie potencjał umożliwiający jej kontrolę nad światem. 22 marca 2023 roku na stronach Future of Life Institute opublikowano list otwarty wzywający wszystkie laboratoria pracujące obecnie nad sztuczną inteligencją do natychmiastowego zatrzymania na co najmniej sześć miesięcy rozwoju i szkolenia systemów AI silniejszych niż Generative Pre-trained Transformer 4 (GPT-4) firmy OpenAI⁵. Pod listem podpisali się między innymi Elon Musk, Yuval Noah Harari, Steve Wozniak, Jaan Tallinn i wielu badaczy z obszaru AI. List wyłuszcza znane z mass mediów (a także z filmów i literatury) obawy przed sztuczną inteligencją wyższą niż ludzka.

¹ N. B o s t r o m, *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press, Oxford 2014.

² M. T e g m a r k, *Life 3.0: Being Human in the Age of Artificial Intelligence*, Knopf Publishing Group, New York 2017.

³ R. K u r z w e i l, *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*, Penguin Books, New York and London 2000.

⁴ R. W a r w i c k, *March of the Machines: The Breakthrough in Artificial Intelligence*, University of Illinois Press, Champaign 2004.

⁵ Zob. *Pause Giant AI Experiments: An Open Letter*, <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>.

„Musimy – głosi list – zadać sobie pytanie: czy powinniśmy pozwolić maszynom zalewać nasze kanały informacyjne propagandą i nieprawdą? Czy powinniśmy zautomatyzować wszystkie zadania, w tym te [nas] spełniające? Czy powinniśmy rozwijać pozaludzkie umysły, które mogą ostatecznie przewyższyć nas liczebnie, przechytryć, uznać za przestarzałych i zastąpić? Czy powinniśmy ryzykować utratę kontroli nad naszą cywilizacją?”⁶. Równie często jak o zagrożeniach słyszemy też o korzyściach płynących z wykorzystania SI, poczynając od szybkiego dostępu do informacji, poprzez wzrost wydajności pracy, bezpieczeństwa (a także innych społecznych wartości, na przykład edukacji czy demokracji), aż po opiekę nad osobami starszymi czy chorymi. Doniesień o kolejnych osiągnięciach jest imponująco dużo. Śledząc nowiny ze świata SI, nieodmiennie oscylujemy między fascynacją i pełnym nadziei oczekiwaniem a niepokojem i lękiem o przyszłość. To nie przypadek, zarówno bowiem obawy, jak i nadzieje są całkowicie uzasadnione. Racja jest fundamentalna: już dziś można dostrzec, jak wiele ważnych zmian dokonało się i wciąż dokonuje praktycznie we wszystkich obszarach codziennego życia ludzi pod wpływem coraz intensywniejszego nasycenia środowiska obiektami wyposażonymi w sztuczną inteligencję. Spodziewamy się zatem, że jeszcze więcej nowości pojawi się nadchodzących latach.

Rozpatrywane są różne prognozy dotyczące miejsca i roli człowieka w świecie opanowanym przez samodoskonalącą się superinteligencję – od wizji optymistycznych, przewidujących, że cały stale rosnący potencjał intelektualny i sprawczy SI wykorzystywany będzie dla dobra ludzkości, aż po skrajnie katastroficzne, zapowiadające zagładę ludzkości i kolonizację całego wszechświata przez uwolnioną z więzów zależności od człowieka nową postać inteligentnego życia⁷. Trudno orzec, która z prognoz jest bliższa spełnienia. Wydaje się jednak wysoce prawdopodobne, że cywilizacja naukowo-techniczna nieodwracalnie weszła już na drogę wiodącą ku autonomicznej technoewolucji (przepowiedzianej już w latach sześćdziesiątych ubiegłego wieku przez Stanisława Lema w dziele *Summa technologiae*⁸), która może wymknąć się spod ludzkiej kontroli. W celu prowadzenia refleksji nad SI powstały między innymi Consortium for the Benevolent Consciousness of Artificial Intelligence, Future of Life Institute i inne podobne organizacje o światowym zasięgu, a Organizacja Narodów Zjednoczonych rozważa powołanie agendy zajmującej się SI na wzór Międzynarodowej Agencji Energii Atomowej⁹.

⁶ Tamże (tłum. fragm. A.L.K., K.W.)

⁷ Por. E. Yudkowsky, *Artificial Intelligence as a Positive and Negative Factor in Global Risk*, w: *Global Catastrophic Risk*, red. N. Bostrom, M. Čirković, Oxford University Press, Oxford 2008, s. 308-362.

⁸ Zob. S. Lem, *Summa technologiae*, Wydawnictwo Literackie, Kraków 1964.

⁹ Zob. M. Druś, *Sekretarz generalny ONZ jest za międzynarodowym nadzorem nad sztuczną inteligencją*, <https://www.pb.pl/sekretarz-generalny-onz-jest-za-miedzynarodowym-nadzorem-nad-sztuczna-inteligencja-1187767>.

Tytułowe pytanie ma właściwie charakter retoryczny. Sztuczna inteligencja będzie trwałym elementem naszego ludzkiego świata, wpływając na kształt życia i wartości. Obecne czasy nazywane są czwartą rewolucją przemysłową. Świadomość, że w dziejach ludzkości i w organizacji światowego ładu nadchodzi wielkie, znaczące zmiany, budzi potrzebę racjonalnego, odpowiedzialnego namysłu nad długofalowym planem działania, wyzwala szeroką skalę emocji, pobudza wyobraźnię. Suma tych trzech elementów, a dokładniej ich wzajemne oddziaływanie i emergentna interferencja, wytwarza cenny kapitał kulturowy, który jako ludzkość powinniśmy jak najlepiej wykorzystać w trosce o lepszą i bezpieczniejszą przyszłość całej planetarnej populacji oraz jej naturalnego i cywilizacyjnego środowiska.

O problemach związanych ze sztuczną inteligencją można myśleć i pisać na wiele sposobów. Można do nich podchodzić w sposób ściśle techniczny, formułując konkretne zadania konstrukcyjne i poszukując środków ich realizacji; można rozpatrywać te problemy w aspektach: ekonomicznym, ekologicznym, prawnym, etycznym, pedagogicznym i innych; można pytać o możliwości zastosowań SI w wielu ważnych dziedzinach życia i pracy: w przemyśle, w nauce, w twórczości artystycznej, w służbie zdrowia, w nauczaniu i wychowaniu czy w wojskowości. Wszystkie te kwestie są ważne i pilne, wymagają wyężonej pracy koncepcyjnej oraz jak najlepszej organizacji i synchronizacji działań. Należy też zwrócić uwagę na dwa obszary wymagające refleksji, które są równie ważne, lecz trudniej uchwytne, bo nie mieszczą się w ramach konkretnych dyscyplin i kompetencji. Pierwszy z nich to troska o szeroko rozumiane cyberbezpieczeństwo. Obejmuje ona między innymi namysł nad sposobami skutecznej ochrony ludzkości przed użyciem zasobów SI w złej wierze i dla osiągnięcia niegodziwych celów – na przykład przez grupy przestępcze czy terrorystyczne, ale także przez osoby i środowiska dążące do zdobycia przewagi nad innymi za pomocą nieetycznej manipulacji technologią SI, poprzez wykorzystanie jej na przykład przeciwko konkurentom lub przeciwnikom w walce o rzadkie zasoby. Inny aspekt cyberbezpieczeństwa, którego także nie można pominąć, to opracowanie zawczasu jak najefektywniejszych środków zaradczych chroniących przed uruchomieniem (świadomym lub mimowolnym) niepożądanych i niebezpiecznych ścieżek rozwoju SI, prowadzących do autonomicznego i niekontrolowanego tworzenia systemów, programów i technologii pośrednio lub bezpośrednio zagrażających ludziom. Wszystkie te zagrożenia są realne i pominięcie ich w publicznej debacie oraz w specjalistycznym dyskursie naukowym byłoby niewybaczalną lekkomyślnością. Drugi obszar wymagający głębokiego namysłu to kwestia relacji między ludźmi a systemami sztucznej inteligencji. Chodzi przede wszystkim o wypracowanie funkcjonalnych i dobrze osadzonych w realiach kulturowych modeli myślenia, zachowań i odniesień do nowo powstającej i niekiedy zaskakującej (pozytyw-

nie lub negatywnie) cywilizacyjnej przestrzeni oddziaływań między ludzką a nie-ludzką inteligencją; modeli pozwalających wzmocnić pozytywne interakcje w obrębie owej przestrzeni oraz w miarę możliwości unikać tych złych, niepokojących i niebezpiecznych.

Im bardziej zmienia się świat wokół nas, wypełniając się w coraz większym stopniu złożonymi, niezrozumiałymi i nieprzewidywalnymi narzędziami, mogącymi służyć człowiekowi i poszerzać jego horyzonty, ale też stawiać przed sobą i realizować własne cele, również cele niezgodne z interesem ludzkości, tym bardziej powinniśmy zadbać o wytworzenie bezpiecznej strefy psychicznego komfortu, opartej na rzetelnej wiedzy, a także na mądrze skonstruowanych tekstach kultury¹⁰, pomagających zwykłemu człowiekowi przezwyciężyć wrażenie obcości czy może nawet niesamowitości SI i poczuć się komfortowo w otoczeniu tak bardzo do nas podobnych i równocześnie tak bardzo odmiennych bytów, wyposażonych w „prawie ludzką” inteligencję, a przy tym znacznie wyprzedzających nas w coraz to nowych obszarach kompetencji. Nie jest to łatwe zadanie, lecz nie wolno zaniedbać jego realizacji, gdyż zaniedbanie to skutkowałoby wyobcowaniem człowieka w świecie, w którym ustalanie zasad i reguł postępowania stopniowo przestanie być wyłączną domeną ludzi. Trzeba również zaakceptować nieuchronny proces metamorfozy ludzkiej tożsamości, wywołany przez implantację zaawansowanych wytworów inteligentnej technologii w obręb ciała ludzkiego, bądź też przez zmiany funkcjonowania kory mózgowej pod wpływem nieprzerwanego kontaktu z elektronicznymi narzędziami komunikacji. Co więcej, nie chodzi tylko o mentalne i behawioralne dostosowywanie się jednostek do nowych aspektów zewnętrznej i wewnętrznej rzeczywistości sztucznej inteligencji. To również wykreowanie nowych ram kulturowych, kodów i idiomów, w których dokona się *sui generis* naturalizacja sztucznej inteligencji. Sam termin „sztuczna” niesie w sobie niepokojącą dwuznaczność. Pierwsze znaczenie słowa „sztuczny” (ang. artificial) odnosi się do artefaktu, przedmiotu niebędącego częścią naturalnego środowiska, lecz wytworzonego za pomocą narzędzi na podstawie uprzedniego projektu. Mamy tu do czynienia z opozycją sztuczny-naturalny, w sensie źródła zaistnienia: zrobione przez człowieka-zrodzone przez przyrodę. To znaczenie pozostaje w mocy, choć coraz więcej elementów naszego środowiska naturalnego to artefakty. Czymże jest na przykład ogród, w którym miejsce każdej rośliny zostało zaplanowane, a te pojawiające się samoistnie są bezwzględnie usuwane? Zapewne naturalnym artefaktem...W słowie „sztuczny” pobrzmiewa także pewna – co najmniej potencjalnie wykluczająca – opozycja: to, co sztuczne, jest nienaturalne, co można niekiedy rozumieć jako „prze-

¹⁰ A. M a j, *Przemiany wiedzy w cyberkulturze: badania nad kulturą, komunikacją, wiedzą i mediami*, Wydawnictwo Uniwersytetu Śląskiego, Katowice 2021.

ciwne naturze”. Komponent emocjonalno-wartościujący pojęcia „sztuczność” jest bardzo często negatywny, co nietrudno zauważyć w wielu kontekstach; choćby wówczas, gdy mówimy o sztuczności czyjegoś zachowania albo gdy dostajemy w prezencie sztuczne kwiaty zamiast prawdziwych. Takie skojarzenia, niekoniecznie uświadomione, mówiąc metaforycznie, obciążają hipotekę sztucznej inteligencji, ponieważ już w punkcie wyjścia ewokują wobec niej nieufność, dystans i rezerwę, jeśli nie wręcz niechęć. Chyba nadszedł czas, by przezwyciężyć tego typu uprzedzenia. Nie oznacza to bynajmniej bezkrytycznej akceptacji wszelkich kierunków rozwoju SI. Chodzi raczej o świadomie krytyczne podejście i możliwie rzetelną analizę ewentualnych zagrożeń, a nie powielanie negatywnych stereotypów. Być może strach przed SI rządzącą światem i eliminującą ludzi jest nieuzasadniony, bo – jak twierdzą Jobst Landgrebe i Barry Smith¹¹ – powstanie tak zwanej silnej sztucznej inteligencji jest matematycznie niemożliwe, a dopiero ona we wszystkich aspektach mogłaby przewyższyć inteligencję ludzką. Nie oznacza to, że ryzyko egzystencjalne związane z wykorzystaniem SI w rozmaitych obszarach życia społecznego nie wymaga namysłu, także etycznego, a nawet regulacji prawnych¹².

Próby odpowiedzialnej refleksji nad wieloma z poruszonych wyżej zagadnień podejmują autorzy tekstów zamieszczonych w niniejszym tomie „Ethosu”. Wiele uwagi poświęcają oni analizie przemian zachodzących w języku, który staje się narzędziem kulturowego „udomowienia” sztucznej inteligencji. Podjęmą też problem funkcji kulturowej i znaczenia tekstów literackich poświęconych relacjom między człowiekiem a SI. Szczególnie pasjonującym wątkiem, który pojawia się w artykułach zamieszczonych w tym tomie, jest pytanie o „etykę maszyn”: czy roboty i inne obiekty SI przyswoją sobie właściwe dla ludzkiej kultury wartości i normy etyczne, czy też stworzą własną moralność, być może pozbawioną humanistycznej wrażliwości?¹³ W przypadku robotów medycznych, którym poświęcony jest jeden z tekstów, pytanie to staje się kluczowe. Równie pasjonującym problemem jest kwestia powiązań między rozwojem SI a imperatywem ochrony środowiska naturalnego: czy sztuczna inteligencja uratuje świat dzięki implementacji optymalnych działań proekologicznych w skali planetarnej, czy przeciwnie – przyspieszy globalną katastrofę?

¹¹ Zob. J. Landgrebe, B. Smith, *Why Machines Will Never Rule the World: Artificial Intelligence without Fear*, Routledge, New York – London 2023.

¹² Do końca 2023 roku Unia Europejska zamierza poddać regulacji prawnej obszar sztucznej inteligencji. Zob. A. Maj, *Unia Europejska coraz bliżej AI Act. Nowe prawo ma uregulować kwestię sztucznej inteligencji*, <https://fakty.tvn24.pl/fakty-o-swiecie/unia-europejska-coraz-blizej-ai-act-nowe-prawo-ma-uregulowac-kwestie-sztucznej-inteligencji-7164166>.

¹³ Por. K. Warwick, *Cyborg Morals, Cyborg Values, Cyborg Ethics*, „Ethics and Information Technology” 5(2003) nr 3, s. 131-137.

Każda inicjatywa zmierzająca do zaangażowania intelektu, uczuć i wyobraźni we wspólną pracę nad powszechnie akceptowalnym projektem optymalnej koegzystencji (kohabitacji?) człowieka z obecnie istniejącą i przyszłą sztuczną inteligencją, powinna zostać przyjęta z życzliwością i satysfakcją. Redaktorzy niniejszego tomu mają nadzieję, że zamieszczone w nim prace autorów pochodzących z różnych środowisk akademickich przyczynią się do pogłębienia i upowszechnienia takiego właśnie – całościowego, integrującego rozmaite punkty widzenia i angażującego wszystkie wymiary ludzkiej percepcji rzeczywistości – sposobu ujmowania problematyki sztucznej inteligencji. Co dalej nastąpi, tego nie da się dokładnie przewidzieć, ale z pewnością wiadomo, że najgorszym z możliwych rozwiązań jest obojętność wobec nadchodzącej przyszłości i bierne czekanie na rozwój wydarzeń. Dopóki mamy jakikolwiek wpływ na kierunki rozwoju SI, musimy uczynić wszystko, by zmaksymalizować szansę na powstanie dobroczynnej sztucznej inteligencji, życzliwej i przyjaznej człowiekowi w rezultacie intensywnych naukowych badań.

*Agnieszka Lekka-Kowalik
Krzysztof Wieczorek*