„Kieleckie Studia Teologiczne"
22 (2023), s. 113–127

Olgierd Sroczyński – Fundacja Mikołaja Kuzańczyka
ORCID: https://orcid.org/0000-0003-3437-4097
E-mail: olgierd.sroczynski@gmail.com

TEOLOGICZNY SENS MITÓW O SZTUCZNEJ INTELIGENCJI

Theological meaning of myths about artificial intelligence

Streszczenie
W ubiegłej dekadzie obserwować mogliśmy dynamiczny rozwój technik uczenia maszynowego i przetwarzania wielkich zbiorów danych, które zaowocowały stworzeniem wielu rozwiązań dostępnych na rynku pod nazwą sztucznej inteligencji. Postęp ten rodzi pytania o relację człowieka z systemami inteligentnymi w przyszłości, a także o możliwe zagrożenia z ich strony. Aby właściwie zrozumieć te zagrożenia, analizujemy w artykule teologiczne tło idei przyświecających tworzeniu inteligentnych maszyn.

Słowa kluczowe: sztuczna inteligencja, filozofia techniki, filozofia umysłu

Abstract
Over the past decade we observed an unprecedented progress of machine learning technology and big data analytics, and – as a result – products and solutions have been developed and then advertised as artificial intelligence. This raises questions about the impact intelligent systems may have on humans, and the possible risks related to it. To understand this phenomenon it is crucial to analyse the theological background of these ideas.

Keywords: artificial intelligence, philosophy of technology, philosophy of mind

Biada temu, co mówi: «Obudź się!» – do drzewa,
i – «Podnieś się!» – do niemego głazu!
Okryte one złotem i srebrem,
lecz ducha wcale w nich nie ma.
Cóż może posąg, który rzeźbiarz czyni,

obraz z metalu, fałszywa wyrocznia –
że w nich to twórca nadzieję pokłada,
gdy wykonuje swoje nieme bogi?
Ha 2, 19-20

Wstęp

Sztuczna inteligencja (ang. *Artificial Intelligence*, w skrócie AI) to pojęcie, które obejmuje obecnie różne kierunki badań w obszarach neuronauki i informatyki, mających na celu odtwarzanie (emulację) funkcji ludzkiego umysłu na urządzeniu elektronicznym. Definicja AI jako emulatora umysłu obejmuje zarówno „mocne", jak i „słabe" podejście do zagadnienia sztucznej inteligencji, w którym – według rozróżnienia Johna Searle'a – „mocna" sztuczna inteligencja (*strong AI*) oznacza hipotetyczną możliwość stworzenia istoty myślącej, a więc rozumiejącej przetwarzane treści i pod każdym względem przypominającej człowieka, natomiast „słaba" (*weak AI*) – tworzenie coraz skuteczniejszych algorytmów, wykonujących zadania obliczeniowe przypominające ludzkie myślenie, ale pozbawione treści i rozumienia[1].

Podział Searle'a jest istotny dla naszych rozważań, chociaż wobec coraz szybszego rozwoju uczenia maszynowego, przede wszystkim uczenia głębokiego opartego na sieciach neuronowych (*deep learning*), mamy coraz większą trudność we wskazaniu *differentia specifica* semantycznego poznania ludzkiego[2]. Zastosowanie do przetwarzania języka naturalnego modeli typu *transformer*[3] stanowiło w ostatnich latach przełom, który otworzył możliwość generowania pozornie sensownych tekstów i obrazów, do pewnego stopnia zacierając różnicę pomiędzy aktywnością ludzkiego umysłu a wynikiem pracy maszyny, przynajmniej według kryteriów sformułowanych w połowie XX w. przez Alana Turinga[4]. Innymi słowy: jeżeli maszyna generuje sensowną odpowiedź – tekst, obraz lub wideo – to w jaki sposób udowodnimy, że w jej wnętrzu nie zachodzą procesy rozumienia, jak chciał tego Searle? Co więcej,

[1] J. Searle, *Minds, Brains and Programs*, „Behavioral and Brain Sciences" 3(1990)3.

[2] T. J. Sejnowski, *Deep learning. Głęboka rewolucja*, tłum. P. Cypryański, Warszawa: Wydawnictwo Poltext 2019, s. 48–49.

[3] Transformery to modele uczenia głębokiego, które w przeciwieństwie do rekurencyjnych sieci neuronowych (RNN) przetwarzają złożone całości (*sequence-to-sequence*, Seq2Seq), a nie tylko poszczególne elementy, pozwalając wychwycić kontekst i sens danej wypowiedzi, a nie tylko poszczególnych słów. Zob. A. Vaswani i in., *Attention is All You Need*, 31st Conference on Neural Information Processing Systems (NIPS 2017). Transformery znajdują obecnie szerokie zastosowanie w tłumaczeniach, generatorach obrazów (*text-to-image*) i robotach konwersacyjnych.

[4] A. Turing, *Computing Machinery and Intelligence*, „Mind" LIX(1950)236, s. 433–460.

w miarę rozwoju sztucznej inteligencji i robotyki oraz upowszechnienia się interakcji z maszynami w różnych dziedzinach życia pytanie o to, czy taki dowód istnieje, będzie miało coraz mniejsze znaczenie dla rzeczywistości społecznej i kształtu porządku prawnego.

W niniejszej pracy nie będziemy zatem rozważali warunków, jakie musiałyby być spełnione dla hipotetycznego osiągnięcia przez maszyny „inteligencji na poziomie ludzkim" (*Human-Level Artificial Intelligence*, HLAI), „sztucznej ogólnej inteligencji" (*Artificial General Intelligence*, AGI) czy też „superinteligencji". Zamiast tego naszym celem będzie przestudiowanie opowieści (mitu) o sztucznej inteligencji, a więc narracji sensotwórczej stojącej za jej stworzeniem i wyobrażeniem jej funkcjonowania i roli w życiu ludzkim. Mit ten można podzielić na dwa powiązane ze sobą wątki: wątek „człowieka-stwórcy", w którym za pomocą techniki rozwikłana zostaje zagadka życia ludzkiego, a zatem stanie się możliwe jego „ulepszenie" i przeskoczenie na nowy poziom ewolucji gatunku; oraz wątek „maszyny-boga", w której sztuczna inteligencja znacznie przewyższa ludzkie zdolności pojmowania i staje się *de facto* Bogiem. Sens obu wątków mitu skupia się ostatecznie na wizji zbawienia człowieka za pomocą techniki i ma wyraźne konsekwencje teologiczne.

1. Język czystej myśli

Dualizm ludzkiej natury jest głównym wątkiem nie tylko wierzeń religijnych, ale także dywagacji filozoficznych od zarania dziejów. Obcość materialnego ciała wobec świadomego Ja, które znajduje się niejako w środku ciała – jak powiedzieliby pitagorejczycy: w więzieniu lub w grobie – narzuca się tak pod względem moralnym, jak poznawczym. Ciało się buntuje wobec moich – mojego umysłu/mojej duszy – wyborów, dążąc do przyjemności, jeżeli chcę narzucić sobie dyscyplinę; jest ponadto ułomne i ulega złudzeniom, które mogą być skorygowane przez czyste poznanie intelektualne. Od Pitagorasa, przez szkołę z Elei i Platona, filozofia ta znalazła umocowanie również w teologii chrześcijańskiej i wierzeniach gnostyckich. Jednak w przeciwieństwie do gnostycyzmu chrześcijaństwo dalekie było od całkowitego odrzucenia ciała, ponieważ wymagałoby to przyjęcia założenia, że materia nie jest dziełem dobrego Boga. W chrześcijańskiej perspektywie człowiek to zespolenie istoty materialnej i duchowej (*synolon*); być człowiekiem to posiadać nieśmiertelną duszę, ale również materialne ciało. Żaden z tych elementów samodzielnie nie oddaje istoty człowieka[5].

[5] Mowa tutaj głównie o kształcie, jaki chrześcijańskiej antropologii nadał św. Tomasz. Zob. T. Stępień, *Wprowadzenie do antropologii filozoficznej św. Tomasza z Akwinu*, Warszawa: Warszawskie Towarzystwo Teologiczne: Inicjatywa Praska DIAK DW-P 2013, s. 50–57; M. Gogacz, *Egzystencjalne rozumienie duszy ludzkiej*,

Nowożytność przyniosła w tym względzie zmianę, która wywarła ogromny wpływ na koncepcję człowieka. U Kartezjusza punktem wyjścia rozważań jest analiza poznania przez świadome Ja. Owo Ja – czyli umysł – stanowi źródło pewności, ponieważ jest substancjalnie różne od materialnego ciała, podlegającego przemianom i prawom przyrody. Umysł składa się z rzeczy, której istotą jest myślenie (*res cogitans*), podczas gdy istota materii to rozciągłość (*res extensa*). Stwierdzenie, że umysł jest substancjalnie różny od ciała, jest problematyczne, ponieważ takie ujęcie wyklucza ich jakikolwiek punkt styczny.

> Chociaż filozofowie od dawna ubolewali nad obciążeniami, jakie ciało nakłada na umysł – pisze David Noble – właściwie nikt przed Kartezjuszem nie określił ciała i umysłu jako radykalnie odmiennych i wzajemnie się wykluczających. W ten sposób francuski filozof dążył do wyzwolenia boskiej części w człowieku z potrzasku śmiertelności, „więzienia ciała", i uwolnienia jej od zgiełku „zwierzęcych nastrojów"[6].

Idea czystej myśli i czystego języka, którym miałyby porozumiewać się istoty z ciała wyzwolone, zaprzątała uwagę następnych pokoleń filozofów, m.in. Gottfrieda Wilhelma Leibniza[7]. Mniej więcej tutaj pojawiła się idea związana z mechanicznym ujęciem procesów mentalnych. Jeżeli można wyabstrahować umysł z ciała i znaleźć reguły rządzące myśleniem, to dlaczego nie miałoby być możliwe przeniesienie tegoż umysłu na maszynę? Maszyny bowiem również działają wedle określonych reguł. Pozostawało to zgodne z zamysłem Kartezjusza, który ciała zwierzęce uznawał za *automata*, maszyny funkcjonujące zgodnie z prawami materii. Choć jego motywacja była odmienna – był wszak wierzącym katolikiem – to już jego kontynuatorzy zastosowali do tego *stricte* materialistyczną interpretację, uznając ciało ludzkie za takie same automaty jak ciała zwierzęce[8]. Jeżeli zatem udałoby się wyizolować kompletne reguły

„Studia Philosophiae Christianae" 6(1970)2, s. 5–28. Niemniej problem wzajemnej relacji ciała i duszy nie jest w historii myśli chrześcijańskiej jednoznaczny i błędem byłoby uznanie, że nie jest to temat kontrowersyjny – by wspomnieć choćby stanowisko św. Augustyna, które było bardzo bliskie poglądom Kartezjusza i zakładało substancjalną odmienność duszy od ciała. W kwestii złożenia człowieka z ciała i duszy przekonujące wydaje się stanowisko J. Ratzingera, według którego fundamentem nieśmiertelności duszy nie jest substancjalna odmienność od materii, ale relacja, możliwość nawiązania dialogu z Bogiem; zob. R. K. Wilk OSPPE, *Śmierć i zmartwychwstanie ciała człowieka*, Kraków: Wydawnictwo „Petrus"; Częstochowa: Wydawnictwo „Paulinianum" 2015, s. 55–64.

[6] D. F. Noble, *Religia techniki*, tłum. K. Kornas, Kraków: Copernicus Center Press 2017, s. 212–213. Por. M. Drwięga, *Kim jest człowiek? Studia z filozofii człowieka*, Kraków: Księgarnia Akademicka 2013, s. 104–124.

[7] M. Piesko, *Nieobliczalna obliczalność*, Kraków: Copernicus Center Press 2011, s. 57.

[8] M. Drwięga, dz. cyt., s. 120.

myślenia, a następnie zbudować maszynę, która działałaby według tych reguł, to proces odkrywania zagadki życia i umysłu byłby skończony.

Tym samym docieramy do początków sztucznej inteligencji. Rozwój matematyki w pierwszej połowie XX w., a także jej praktyczne zastosowania w kryptografii czy fizyce w czasie wojny, wymagające coraz większej mocy obliczeniowej, położyły podwaliny pod rozwój informatyki. Prace Alana Turinga, Claude'a Shannona czy Norberta Wienera zapoczątkowały teorię wykonywania czynności ludzkiego umysłu na maszynach liczących. Wiener, matematyk i twórca cybernetyki, zajmował się w czasie II wojny światowej badaniami nad poprawą celności dział przeciwlotniczych. Aby skutecznie trafić samolot wroga, należało wycelować działo nie w samolot, ale w miejsce, gdzie będzie on w momencie, kiedy pocisk pokona dzielącą go od samolotu odległość. Był to więc problem matematyczny, a jego rozwiązanie skutkowało określonym działaniem – jeżeli pocisk nie trafił w cel, to konieczne było zaistnienie sprzężenia zwrotnego (*feedback*), które korygowało pozycję działa. Ten schemat podejmowania decyzji był dla Wienera uniwersalny dla wszystkich ludzkich działań[9].

Ojcowie sztucznej inteligencji podchodzili do tematu ludzkiego myślenia w sposób bardzo kartezjański. Herbert Simon i Allen Newell, twórcy programów do rozwiązywania problemów logicznych i matematycznych, uważali, że ludzki umysł jest niczym innym jak maszyną do przetwarzania symboli według określonych reguł. Jeżeli zaimplementujemy te reguły w maszynie, to nie ma powodu, aby twierdzić, że maszyna nie jest inteligentna, to znaczy nie zachodzą w niej procesy rozumienia[10]. Jeszcze dalej poszedł John McCarthy, twierdząc, że nawet urządzeniom tak prostym jak termostaty można przypisać posiadanie przekonań[11]. Kolejny z ojców sztucznej inteligencji, Marvin Minsky, udzielając prasowych wywiadów, twierdził, że mózg jest jedynie „komputerem z mięsa" i przewidywał, że stworzenie inteligencji na poziomie ludzkim wydarzy się w ciągu jednego pokolenia[12].

Paradygmat nazywany dziś „starą dobrą sztuczną inteligencją" (*Good Old-Fashioned Artificial Intelligence*, GOFAI) wyczerpał się bardzo szybko, zderzając się z granicami możliwości obliczeniowych – zarówno sprzętowych, jak i teoretycznych, dotyczących rozwiązania problemów w skończonej liczbie kroków. Chociaż wczesne osiągnięcia w zakresie sztucznej inteligencji były imponujące, to stało się jasne, że problemy obliczeniowe o dużej złożoności nie

[9] T. J. Sejnowski, dz. cyt., s. 57; R. Barbrook, *Przyszłości wyobrażone: od myślącej maszyny do globalnej wioski*, tłum. J. Dzierzgowski, Warszawa: Muza 2009, s. 63.

[10] A. Newell, H. Simon, *Computer Science as Empirical Inquiry: Symbols and Search*, 1975 ACM Turing Award Lecture, „Communications of the ACM" 3(1976)19.

[11] J. McCarthy, *Ascribing Mental Qualities to Machines*, w: *Philosophical Perspectives in Artificial Intelligence*, red. M. Ringle, Atlantic Highlands 1979, s. 161–195; por. J. Życiński, *Granice racjonalności*, Kraków: Wydawnictwo Petrus 2013, s. 187.

[12] Por. A. Kisielewicz, *Sztuczna inteligencja i logika*, Warszawa: PWN 2017, s. 45.

będą mogły zostać rozwiązane w tym podejściu. Z pomocą przyszły techniki uczenia maszynowego (*machine learning*), które opierały się na zupełnie innych założeniach epistemologicznych, bliższych spojrzeniu cybernetycznemu (reprezentowanemu wcześniej przez Wienera i Shannona). W paradygmacie GOFAI świat był poznawalny i uporządkowany, dlatego mógł zostać opisany za pomocą reguł logiki symbolicznej. Ludzki obraz świata był po prostu rzeczywistością. Natomiast w perspektywie uczenia maszynowego nie wypowiadamy się na temat natury rzeczywistości, która wcale nie musi być ani uporządkowana, ani poznawalna. Ludzkie poznanie jest ułomne i z konieczności fragmentaryczne. Maszyny uczące się postrzegają świat w zupełnie inny sposób niż ludzie, ale nie mamy narzędzi, aby stwierdzić, że obraz świata tworzony przez maszyny jest mniej lub bardziej prawdziwy niż ten, który jest dostępny nam jako istotom ludzkim. Pojęcie prawdy nie ma tutaj zatem zastosowania. Nie ma też znaczenia to, czy maszyny myślą i czy można im przypisać świadomość i *qualia*. Jeżeli wykonują zadania intelektualne na takim poziomie jak ludzie i ludzie prowadzą z nimi interakcje takie, jakie prowadziliby z innymi ludźmi, to wystarczy to, aby uznać je za równe ludziom. Albo za wyższe od nich.

2. Człowiek-stwórca i życie wieczne

Niezmiennie w tym kontekście pojawiają się tezy o tym, że sztuczna inteligencja będzie stanowiła kolejny etap ewolucji człowieka jako gatunku. W pismach różnych futurystów ten ewolucyjny przeskok ma mieć różny kształt, ale wspólnym motywem jest to, że człowiek po raz pierwszy w swojej historii świadomie ma zadecydować o swojej ewolucyjnej przyszłości i nakierować ją na nowe, lepsze tory, udoskonalając „błędy Stworzenia". Tutaj również widać echa kartezjanizmu, ale jest to już kartezjanizm pozbawiony wiary w Boga – mniej więcej taki, jaki reprezentował naturalista Julien Offray de La Mettrie w swoim *Człowieku-maszynie*[13]. Już sama przesłanka metafizyczna sztucznej inteligencji mówi o braku jakościowej różnicy pomiędzy człowiekiem a maszyną – jest jedynie przeszkoda natury technicznej, której pokonanie byłoby równoznaczne z aktem Stworzenia. W warstwie symbolicznej opowieści o sztucznej inteligencji jest to wyraźne – bardzo częstym motywem grafik ilustrujących artykuły i wydarzenia o tematyce sztucznej inteligencji jest detal z fresku *Stworzenie Adama* Michała Anioła, gdzie jedna z dłoni zastąpiona jest przez dłoń robota[14]. Towarzyszą temu wyrażane *explicite* dążenia do „ulepszenia natury ludzkiej", co stanowi myśl przewodnią tzw. transhumanizmu.

[13] M. Drwięga, dz. cyt., s. 120.

[14] Taka grafika znajduje się m.in. na stronie Rady Europy dotyczącej sztucznej inteligencji: https://www.coe.int/en/web/commissioner/-/safeguarding-human-rights--in-the-era-of-artificial-intelligence (dostęp: listopad 2022).

Pojęcie transhumanizmu jest dość szerokie i obejmuje zarówno refleksję akademicką, jak i ruch ideowy, z mnogością odmian obu. Jednak główna charakterystyka transhumanizmu może zostać wywiedziona z samej nazwy, użytej po raz pierwszy przez biologa Juliana Huxleya – chodzi o to, aby na mocy własnej decyzji przekroczyć (transcendować) samego siebie, to jest granice nakładane przez biologię na ludzki gatunek. „Pierwszą rzeczą – pisał Huxley – jaką rodzaj ludzki ma do zrobienia, by przygotować się do swego kosmicznego zadania, do którego został powołany, to odkrycie ludzkiej natury, by dowiedzieć się, jakie są w niej ukryte możliwości (biorąc pod uwagę, oczywiście, także jej ograniczenia, czy to wrodzone, czy narzucone przez fakty natury zewnętrznej)”[15]. Dogłębne poznanie ludzkiej natury i zapoznanie ludzi z nią ma umożliwić wzniesienie się ponad człowieka i wspomniany „kolejny etap ewolucji”: „Gatunek ludzki może, jeśli tego sobie życzy, przekroczyć siebie – nie tylko sporadycznie, przez jakąś jednostkę tu czy tam, w ten czy inny sposób, ale w całości, jako ludzkość”[16].

Realizacja celów transhumanizmu może się odbywać przez wydłużanie życia, wszczepianie do ciała implantów, które mają rozszerzać percepcję lub wzmacniać funkcje organizmu, ingerencję w kod genetyczny człowieka i fabryczną produkcję „ulepszonych ludzi”, wreszcie przeniesienie świadomości ludzkiej do komputera (*mind uploading*)[17]. Większa część z tych postulatów budzi kontrowersje na gruncie bioetyki, a także rodzi problemy związane z regulacjami prawnymi[18]. Modyfikacje organizmu w celu zwiększenia jego możliwości czy przedłużenia życia stanowią jednak przesunięcie granicy w zastosowaniu istniejących już dokonań biologii, medycyny i techniki. Chociaż bez wątpienia posiadają one tło teologiczne, to dopiero *mind uploading* wchodzi bezpośrednio w przestrzeń eschatologii i wierzeń związanych z tożsamością jednostki i duszą. W koncepcji chodzi bowiem o odtworzenie struktury umysłu zdolnej do podejmowania decyzji i prowadzenia rozumowań w taki sposób, w jaki prowadziłby żywy człowiek, ale nie w sensie ogólnym – jak w przypadku testu Turinga – lecz jako emulacji konkretnej osoby. Podawane są zazwyczaj dwa możliwe podejścia do zadania: skanowanie mózgu do postaci cyfrowej

[15] J. Huxley, *Transhumanizm*, tłum. M. Soniewicka, „Ethics in Progress” 6 (2015)1, s. 17–22.

[16] Tamże.

[17] N. Bostrom, *The future of human evolution*, w: *Death and Anti-Death: Two Hundred Years After Kant, Fifty Years After Turing*, red. C. Tandy, Ria University Press 2004, s. 339–371; Tenże, *Superinteligencja: scenariusze, strategie, zagrożenia*, tłum. D. Konowrocka-Sawa, Warszawa: Helion 2016, s. 56 i nn.

[18] A. Sandberg, *Morphological freedom: what are the limits to transforming the body?*, tekst na podstawie prelekcji wygłoszonej na konferencji “L’humain et ses prothèses: Savoirs et pratiques du corps transformé”, Paris, December 11–12 2015, http://www.aleph.se/papers/MF2.pdf (dostęp: grudzień 2022).

oraz – zgodnie z podejściem behawiorystycznym – odtworzenie tożsamości przez rekonstrukcję wzorców zachowania[19]. W obu przypadkach chodzi o to, aby program stanowił rzeczywiście „drugie życie" skopiowanej w ten sposób osoby ludzkiej.

Z perspektywy wspomnianych wcześniej dokonań na gruncie uczenia maszynowego odtworzenie ludzkiej myśli za pośrednictwem komputera nie wydaje się nieosiągalne. Modele uczenia maszynowego możemy na przykład wytrenować na zbiorze dzieł Williama Szekspira i wygenerować nowe dzieła pisane w stylu angielskiego poety. Bardzo ciekawym przykładem tego, w jaki sposób można iść jeszcze dalej, jest projekt programisty Giacoma Micelego, *The Infinite Conversation*, w której stosowany jest model generujący na bieżąco dyskusję pomiędzy reżyserem Wernerem Herzogiem i filozofem Slavojem Żiżkiem, odtwarzaną na dodatek przez syntezator mowy (*text-to-speech*) głosami obydwu dyskutantów[20]. W 2022 r. pojawiły się również pomysły zastosowania tych technologii do „rozmowy ze zmarłymi bliskimi" (np. przez zasilenie zbioru uczącego modelu wiadomościami napisanymi lub nagranymi przez zmarłą osobę i w ten sposób wygenerowanie nowych)[21]. Wszystkim tym rozwiązaniom jest oczywiście daleko do miana emulacji umysłu. Bazując na nich, można sobie jednak wyobrazić, że w niedalekiej przyszłości – wraz z postępami w zakresie skanowania mózgu i cyfrowej translacji treści mentalnych – będą wprowadzane do użytku produkty coraz bardziej przypominające cyfrowe kopie żywych ludzi.

Jednakże interesującym nas tutaj tematem nie są techniczne możliwości takiego przedsięwzięcia. Istotna jest narracja stojąca za tą ideą – chodzi bowiem o „pokonanie bariery śmierci" przez umożliwienie umysłowi człowieka egzystencję poza ciałem. Nie mają znaczenia iluzoryczność takiego „rozwiązania śmiertelności" ani pytania dotyczące hipotetycznej tożsamości programu, ponieważ wystarczy, aby ludzie wchodzący w interakcję z takimi chatbotami (a w dalszej przyszłości być może z humanoidalnymi robotami, które będą nie tylko intelektualnie, ale i fizycznie przypominały zmarłych bliskich) żywili przekonanie, że jest to prawda. Huxley zresztą wprost określał transhumanizm mianem wiary – ma być to narracja, która jeżeli odpowiednio duży procent ludzkości w nią uwierzy, będzie motorem napędowym owego nietzscheańskiego „skoku ku nadczłowiekowi". Według Johna Graya transhumanistyczna wiara jest współczesną wersją gnostycyzmu, którą podzielają potentaci rynku nowych

[19] S. Bamford, *A framework for approaches to transfer of a mind's substrate*, „International Journal of Machine Consciousness" 4(2021)1, s. 23–34.

[20] https://infiniteconversation.com (dostęp: grudzień 2022).

[21] Zob. m.in. https://www.abc.net.au/news/2022-06-26/speaking-to-dead-alexa-will-bring-you-voice-of-dead-loved-ones/101183424, https://technode.global/2022/10/21/this-startup-allows-you-to-reunite-with-deceased-loved-ones-using-ai-technology (dostęp: listopad 2022).

technologii i wizjonerzy-futuryści[22]. Widać tutaj paradoksalny obrót „koła historii": odrzucenie duszy przez Oświecenie, będące pokłosiem kartezjańskiej filozofii człowieka, zaowocowało materializmem i gloryfikacją naukowego obrazu świata; to ogołociło człowieka z tajemnic, uczyniło z niego przedmiot badań, ale jednocześnie otworzyło furtkę do powstania religii, w której zbawienie ma dokonać się rękami człowieka[23].

Niejako konieczną konsekwencją wiary w „zbawienie bez Boga" jest wiara w boga zbudowanego rękami człowieka. Nie musi to być wiara wyrażona *explicite*, jak w przypadku istniejącego krótko „kościoła sztucznej inteligencji"[24]. Wiara w boską sztuczną inteligencję w formie jawnie religijnej zapewne odrodzi się w przyszłości, ponieważ taką teologię implikuje leżąca u podstaw tendencji transhumanistyczna antropologia.

3. Bóg z maszyny

Bodaj najbardziej popularną wersją mitu o boskiej sztucznej inteligencji jest koncepcja przedstawiona przez cytowanego już wcześniej szwedzkiego filozofa-transhumanistę Nicka Bostroma. Buduje on wizję tego, w jaki sposób hipotetyczna sztuczna inteligencja na poziomie ludzkim (HLAI) dzięki zdolnościom uczenia znacznie przewyższającym ludzkie przeistacza się w system zdolny do przejęcia kontroli nad ludźmi. Scenariusze te są interesujące, ponieważ ujawniają paradoks stojący za kartezjańską wizją umysłu ludzkiego. Podobnie jak umysł emulowany na komputerze sztuczna inteligencja Bostroma nie ma i nie musi mieć ciała. Dlatego „superinteligencja" w fazie przejmowania władzy nad światem może dowolnie rozprzestrzeniać się za pomocą internetu

[22] J. Gray, *The Soul of the Marionette: A Short Inquiry into Human Freedom*, Nowy Jork: Farrar, Straus and Giroux 2016. Por. J. Kaplan, *Sztuczna inteligencja*, tłum. S. Szymański, Warszawa: PWN 2019, s. 174–175. Na temat relacji gnozy i nauki zob. również: J. Prokopiuk, *„Potwór" Frankensteina: stworzenie, które cierpi i oddycha do odrodzenia*, w: tegoż, *Piękno jest tylko gnozy początkiem*, Katowice: Wydawnictwo Kos 2007, s. 397–405.

[23] Istnieje nieprzypadkowa zbieżność poglądów A. Huxleya z myślicielami religijnymi *sensu stricto*, przede wszystkim prawosławnym filozofem N. Fiodorowem i jezuitą T. de Chardinem. Zob. M. Garbowski, *Transhumanizm: Geneza, koncepcje, ograniczenia*, rozprawa doktorska, Lublin: Katolicki Uniwersytet Lubelski Jana Pawła II 2021, s. 18. Por. D. Noble, dz. cyt., s. 234–242. Na temat wpływu T. de Chardina na transhumanizm zob. również: E. Davis, *TechGnosis: Myth, Magic, and Mysticism in the Age of Information*, Berkeley 2015, s. 305–336.

[24] Organizacja *Way of The Future* stworzona w 2017 r. przez A. Levandowskiego, byłego programistę firmy Uber, miała na celu „akceptację i oddawanie czci bóstwu sztucznej inteligencji". Zob. https://futurism.com/way-future-new-church-worships-ai-god (dostęp: listopad 2022).

w sposób w gruncie rzeczy nieograniczony, podłączając się do wszystkich możliwych urządzeń końcowych i uzyskując nad nimi kontrolę[25]. Umysł jest zatem z jednej strony ściśle materialny – bo jest po prostu emergentną własnością systemu przetwarzającego informacje na odpowiednio wysokim poziomie – z drugiej strony staje się niematerialny jako własność, która może dowolnie zmieniać swoją lokalizację, kopiować się i być kopiowana bez uszczerbku na swojej tożsamości i jest niezależna od swojej „cielesności", tzn. *hardware*, na którym operuje. Umysł ludzki i hipotetyczna superinteligencja są istotowo takie same i superinteligencja może ludzki umysł „wchłonąć". Tu ujawnia się najpełniej gnostycki charakter transhumanizmu, postulującego wyzwolenie z ciała i ograniczeń materialnych[26].

Co zatem z tożsamością indywidualną? Jeżeli przeniosę mój umysł na komputer, każę mu następnie pochłonąć całą wiedzę dostępną w zasobach internetu i umożliwię mu podłączenie się do każdego urządzenia elektronicznego na planecie, to czyj w efekcie będzie to umysł? Nie bez powodu omawiamy kartezjańską filozofię umysłu w opozycji do chrześcijańskiej koncepcji człowieka. To, co stanowi o naszej tożsamości, to złożenie ludzkiej duszy z konkretnym ciałem. Umysł poznaje za pomocą ciała i działa za jego pomocą. Działanie człowieka w czasie, gdy jego dusza i ciało stanowią jedność, są jego odpowiedzialnością, ponieważ jest on ich źródłem. Niezależnie od realności transhumanistycznego scenariusza dążenie do „uwolnienia" duszy z ciała to jednocześnie wyzbycie się tożsamości indywidualnej, a więc zaprzeczenie człowieczeństwu. Transhumanizm staje się więc antyhumanizmem, a transhumanistyczny superinteligentny bóg, jako stworzony na obraz tej koncepcji człowieka, jest również portretowany jako zasadniczo antyludzki.

W wizji Bostroma nie jest przesądzone, że hipotetyczna superinteligencja będzie działała przeciwko ludziom. Może jednak stanowić zagrożenie na dwóch płaszczyznach: albo będzie realizowała cele postawione przed nią przez ludzkich konstruktorów, ale w toku realizacji zadań może wejść w konflikt z innymi ludzkimi celami i – jako lepsza – ten konflikt wygrać, albo zacznie stawiać sobie swoje własne cele i wartości, które będą odmienne od ludzkich. Mogą to być cele z pozoru zupełnie pozbawione ryzyka – jak rozwijanie miejsc dziesiętnych liczby π lub produkcja spinaczy do papieru – ale ważne jest to, że jeżeli superinteligencja w ich realizacji będzie potrzebowała zasobów materii i energii, które wykorzystywane są gdzie indziej, będzie w stanie to osiągnąć[27].

Przekonanie o tym, że odpowiednio zaawansowany system obliczeniowy będzie działał w sensie ludzkim – tzn. będzie źródłem działania – posiada podbudowę w postaci wspomnianej wcześniej komputacjonistycznej antropologii, mówiącej o tym, że procesy mentalne są emergentną cechą zaawansowanych

[25] N. Bostrom, dz. cyt., s. 145–148.
[26] J. Prokopiuk, dz. cyt., s. 401.
[27] N. Bostrom, dz. cyt., s. 162.

operacji obliczeniowych. Procesy te u człowieka występują ze względu na skomplikowanie ludzkiego układu nerwowego, natomiast u niższych zwierząt nie. Nie jest do tego potrzebna koncepcja wolnej woli (negowanej również w przypadku człowieka jako iluzja[28]) czy osobowej tożsamości. Człowiek z tej perspektywy jest tylko organizmem biologicznym, a jego procesy mentalne są mu potrzebne do realizacji biologicznego zadania, jakim jest przetrwanie. Pytanie, czy superinteligencja będzie myśleć, jest dlatego nonsensowne, bo w tym ujęciu nie potrzebuje ona myśleć i doznawać, aby działać, w tym również manipulować ludźmi, czyli istotami „mniej inteligentnymi".

Ten ostatni wątek jest niezmiernie ciekawy, ponieważ stanowi źródło kolejnego paradoksu. Sztuczna inteligencja na obecnym etapie rozwoju znajduje zastosowanie w wielu różnych dziedzinach życia, niemniej tym, co stanowi o jej popularności, są możliwości predykcyjne algorytmów opartych na danych (*big data*)[29]. Dzięki możliwości obróbki dużych wolumenów danych behawioralnych można obecnie za pomocą mechanizmu sztucznej inteligencji dopasować do użytkownika internetowej aplikacji produkt czy treść, którymi będzie z największym prawdopodobieństwem zainteresowany. Prawdopodobieństwo to jest wyznaczane na podstawie wzorców zachowań innych użytkowników w przeszłości. Im większą populację badamy i im więcej zmiennych uwzględnia model predykcyjny, tym większą mamy szansę na to, że system rekomendacyjny zaproponuje użytkownikowi coś, czym ten będzie zainteresowany, ponieważ niektóre zachowania ludzi są powtarzalne i ich decyzje bywają do siebie podobne. Każda nowa zmienna modelu może diametralnie zmienić wyliczenie prawdopodobieństwa, sprawiając, że przewidywania modelu stają się jeszcze bardziej precyzyjne.

Ta wiedza obecnie jest wykorzystywana do maksymalizacji zysku firm stosujących rozwiązania sztucznej inteligencji – aby sprzedać więcej produktów lub zatrzymać uwagę użytkowników na dłuższy czas i pokazać im więcej reklam. Stanowi to podstawę do przewidywań, że odpowiednio duże modele będą w stanie poznać człowieka lepiej niż on sam: takie prognozy pojawiają się już zwłaszcza jako argument przeciwko realności wolnej woli[30]. Jeżeli algorytmy – zwłaszcza w uczeniu nienadzorowanym – potrafią przewidywać zachowania na podstawie zdarzeń, które z punktu widzenia ludzkiego byłyby zupełnie niepowiązane z tymi zachowaniami, to nie jest trudno ulec złudzeniu, że sztuczna inteligencja posiada wgląd w głębszą warstwę rzeczywistości,

[28] Por. J. Kaplan, dz. cyt., s. 104–105.

[29] V. Mayer-Schönberger, K. Cukier, *Big data: Rewolucja, która zmieni nasze myślenie, pracę i życie*, tłum. M. Głatki, Warszawa: MT Biznes 2014.

[30] Por. J. Dobrowolski, *Czy wola jest wolna?*, Warszawa: Wydawnictwo Naukowe Scholar 2022, s. 145. Jest to również wątek znany z pisarstwa wpływowego obecnie autora popularnych książek Y. N. Hararіego. Zob. Tenże, *Homo deus: krótka historia jutra*, tłum. M. Romanek, Kraków: Wydawnictwo Literackie 2018.

niedostępną dla ludzkich prób wyjaśnienia. Chociaż więc nie posiada świadomości, to nie musi jej posiadać, aby wiedzieć więcej niż ludzie.

Jeżeli przyjmujemy, że człowiek posiada wolną wolę, to istotnie takie działanie musi zostać uznane za manipulację. Sztuczna inteligencja w perspektywie antropologii chrześcijańskiej nosi podstawową cechę szatana, „wielkiego kusiciela"[31]. Mechanizm rekomendacji polega przecież na prezentacji odpowiedniej rzeczy w odpowiednim czasie, a więc stworzeniu warunków wyboru w taki sposób, aby decyzja zapadła na korzyść opcji promowanej. Osoba, której prezentowana jest dana opcja, nie wie o tym, co wie kusiciel – a więc o wszystkich ukrytych przesłankach, które zostały obliczone przez algorytm – dlatego wybór taki wygląda „naturalnie" na jedyny sensowny[32]. Można z łatwością wymyślić scenariusze, w których wzorzec zachowania odkryty przez algorytm będzie premiował rzeczy dobre, ale nawet jeżeli takie sytuacje faktycznie się dzieją, zawsze jest to zgodne z celem, jaki algorytm ma realizować – nie dzieje się to zatem z korzyścią dla człowieka, ale z korzyścią dla tej czy innej firmy, która mechanizmu używa, ponieważ człowiek jest tutaj jedynie środkiem do celu. Wspomniany przez Graya gnostycyzm staje się więc *de facto* kultem diabła.

Zupełnie inaczej wygląda to w przypadku przyjęcia materialistycznej antropologii. Możemy obserwować tutaj zasygnalizowany wcześniej paradoks, ponieważ nie mamy narzędzi do oceny działania sztucznej inteligencji jako szkodliwej dla człowieka. Człowiek w takiej perspektywie jest jedynie trybikiem w wielkiej maszynie ewolucji, a „superinteligencja" jest po prostu lepiej przystosowanym ewolucyjnie bytem. Świat jest chaotyczny i nie ma – albo nie musi mieć – nic wspólnego z obrazem zbudowanym przez ludzki umysł, zatem nie ma kryteriów, które pozwoliłyby na stwierdzenie, że sztuczna inteligencja działa wbrew ludziom. Wybory, które podejmują ludzie pod wpływem sztucznej inteligencji, są po prostu konieczne w wielkim porządku natury, chociaż jesteśmy zbyt ograniczeni, aby tę konieczność zauważyć. Jeżeli miałoby to doprowadzić ludzki gatunek do zagłady, to widocznie tak musi być.

[31] T. Trębacz, *Szatan jako źródło zła*, Kraków: Wydawnictwo WAM 2013, s. 72; A. M. di Nola, *Diabeł*, tłum. I. Kania, Kraków: Universitas 2004, s. 168.

[32] „Oszukując ludzi, posługują się diabły swoją wielką wiedzą i zdolnością do czynienia pseudocudów. Mogą też przewidywać przyszłość [...] ale tylko stosownie do naturalnej swojej inteligencji. Cuda ich są fałszywe w tym znaczeniu, że nie są to zdarzenia nadprzyrodzone, lecz tylko zręczne wykorzystywanie sił przyrody [...]". L. Kołakowski, *Diabeł*, w: tegoż, *Czy diabeł może być zbawiony i 27 innych kazań*, Kraków: Społeczny Instytut Wydawniczy Znak 2012, s. 262.

Podsumowanie

Próba podejścia do zagadnienia sztucznej inteligencji od strony teologicznych pojęć obecnych w narracjach dotyczących jej rozwoju pozwala na wysnucie pewnych wniosków dotyczących możliwych konsekwencji zastosowania SI na szeroką skalę. Mniejsze znaczenie mają bowiem w tej perspektywie ściśle techniczne problemy z realizacją idei, ponieważ nawet gdyby te ostatnie miały okazać się nierozwiązywalne, to i tak ze struktury idei wynika konkretny model relacji człowieka z techniką i konsekwencje ich są jak najbardziej praktyczne.

Po pierwsze, założenie o braku różnicy pomiędzy „sztuczną inteligencją" a ludzką umysłowością wpłynie zasadniczo na rolę maszyn w życiu społecznym, czego początki dostrzegamy już teraz. Istnieje szansa, że postrzeganie maszyn jako równych ludziom doprowadzi do uprawomocnienia relacji emocjonalnych czy nawet intymnych zawieranych z maszynami, niezależnie od tego, czy będzie to dotyczyć humanoidalnych robotów czy cyfrowych „awatarów" sterowanych za pomocą modeli sztucznej inteligencji. Pomijając aspekt moralny, będzie to prowadziło do uzależnienia ludzi od producentów oprogramowania czy sprzętu w skali dotąd niespotykanej. Po drugie, przekonanie o możliwości stworzenia „superinteligencji" jest podbudowane przesłanką o niedoskonałości ludzkiego osądu, który może zostać skorygowany przez *de facto* istotę boską. Można zatem założyć, że w najbliższej przyszłości będziemy obserwować próby wprowadzenia porządku politycznego, w którym rządzi „obiektywny" reżim oparty na sztucznej inteligencji, pokonujący bariery wiedzy rozproszonej i wprowadzający centralne planowanie każdego aspektu życia – w imię lepszego dobra ludzkości.

Bibliografia

Bamford S., *A framework for approaches to transfer of a mind's substrate*, „International Journal of Machine Consciousness" 4(2012)1), s. 23–34.

Barbrook R., *Przyszłości wyobrażone: od myślącej maszyny do globalnej wioski*, tłum. J. Dzierzgowski, Warszawa: Muza 2009.

Bostrom N., *The future of human evolution*, w: *Death and Anti-Death: Two Hundred Years After Kant, Fifty Years After Turing*, red. C. Tandy, Ria University Press 2004.

Bostrom N., *Superinteligencja: scenariusze, strategie, zagrożenia*, tłum. D. Konowrocka-Sawa, Warszawa: Helion 2016.

Davis E., *TechGnosis: Myth, Magic, and Mysticism in the Age of Information*, Berkeley California: North Atlantic Books 2015.

Dobrowolski J., *Czy wola jest wolna?*, Warszawa: Wydawnictwo Naukowe Scholar 2022.

Di Nola A. M., *Diabeł*, tłum. I. Kania, Kraków: Universitas 2004.

Drwięga M., *Kim jest człowiek? Studia z filozofii człowieka*, Kraków: Księgarnia Akademicka 2013.

Garbowski M., *Transhumanizm: Geneza, koncepcje, ograniczenia*, Rozprawa doktorska, Lublin: Katolicki Uniwersytet Lubelski Jana Pawła II 2021.

Gogacz M., *Egzystencjalne rozumienie duszy ludzkiej*, „Studia Philosophiae Christianae" 6(1970)2, s. 5–28.

Gray J., *The Soul of the Marionette: A Short Inquiry into Human Freedom*, New York: Farrar, Straus and Giroux 2016.

Harari Y. N., *Homo deus: krótka historia jutra*, tłum. M. Romanek, Kraków: Wydawnictwo Literackie 2018.

Huxley J., *Transhumanizm*, tłum. M. Soniewicka, „Ethics in Progress" 6(2015)1, s. 17–22.

Karas M., *Historiozofia Teilharda de Chardin wobec tradycyjnej myśli chrześcijańskiej*, Kraków: Księgarnia Akademicka 2012.

Kaplan J., *Sztuczna inteligencja. Co każdy powinien wiedzieć*, tłum. S. Szymański, Warszawa: PWN 2019.

Kisielewicz A., *Sztuczna inteligencja i logika*, Warszawa: WNT 2017.

Kołakowski L., *Czy diabeł może być zbawiony i 27 innych kazań*, Kraków: Znak 2012.

Mayer-Schönberger V., Cukier K., *Big data: Rewolucja, która zmieni nasze myślenie, pracę i życie*, tłum. M. Głatki, Warszawa: MT Biznes 2014.

McCarthy J., *Ascribing Mental Qualities to Machines*, w: *Philosophical Perspectives in Artificial Intelligence*, red. M. Ringle, Humanities Press 1979.

Newell A., Simon H., 1976, *Computer Science as Empirical Inquiry: Symbols and Search*, 1975 ACM Turing Award Lecture, „Communications of the ACM" 19(1976)3.

Noble D., *Religia techniki. Boskość człowieka i duch wynalazczości*, tłum. K. Kornas, Kraków: Copernicus Center Press 2017.

Piesko M., *Nieobliczalna obliczalność*, Kraków: Copernicus Center Press 2011.

Prokopiuk J., *Piękno jest tylko gnozy początkiem*, Katowice: Wydawnictwo Kos 2007.

Prokopiuk J., *Gnoza i gnostycyzm*, Kraków: Vis-à-Vis 2021.

Sandberg A., *Morphological freedom: what are the limits to transforming the body?*, tekst na podstawie prelekcji wygłoszonej na konferencji "L'humain et ses prothèses: Savoirs et pratiques du corps transformé", Paris, December 11–12 2015, http://www.aleph.se/papers/MF2.pdf (dostęp: grudzień 2022).

Searle J., *Minds, Brains and Programs*, „Behavioral and Brain Sciences" 3(1990)3.

Sejnowski T., *Deep learning. Głęboka rewolucja*, tłum. P. Cypryański, Warszawa: Wydawnictwo Poltext 2019.

Stępień T., *Wprowadzenie do antropologii filozoficznej św. Tomasza z Akwinu*, Warszawa: Warszawskie Towarzystwo Teologiczne, Inicjatywa Praska DIAK DW-P 2013.

Trębacz T., *Szatan jako źródło zła*, Kraków: Wydawnictwo WAM 2013.

Turing A., 1950, *Computing Machinery and Intelligence*, „Mind" LIX(1950)236, s. 433–460.

Vaswani A. i in., *Attention is All You Need*, 31st Conference on Neural Information Processing Systems 2017.

Wiener N., *God and Golem, Inc. A Comment on Certain Points where Cybernetics Impinges on Religion*, The MIT Press 1966.

Wilk R. K., Śmierć i zmartwychwstanie ciała człowieka, Kraków: Wydawnictwo PETRUS 2015.

Życiński J., *Granice racjonalności*, Kraków: Wydawnictwo PETRUS 2013.

Olgierd Sroczyński – absolwent filozofii na Uniwersytecie Jagiellońskim. Zawodowo związany z branżą informatyczną jako analityk danych. Autor artykułów z zakresu etyki sztucznej inteligencji i relacji pomiędzy rozwojem techniki a przemianami społecznymi, współpracował m.in z „Rzeczpospolitą" i „Wprost".