Rocznik Filozoficzny Ignatianum
The Ignatianum Philosophical Yearbook
Vol. 30, No. 1 (2024) s. 225–234
PL ISSN 2300–1402
MATERIAŁY I RECENZJE

Damian Szczęch
ORCID: 0000-0003-2290-1726
Katolicki Uniwersytet Lubelski Jana Pawła II

Jobst Landgrebe, Barry Smith, *Why Machines Will Never Rule The World. Artificial Intelligence without fear* (New York–London: Routledge, 2023), ss. 354. ISBN: 978-1-003-31010-5. DOI: 10.4324/9781003310105

Książka składa się z przedmowy (*foreword*), wstępu, 13 rozdziałów (podzielonych na trzy części), załącznika zawierającego krótki opis teorii turbulencji, słownika (*glossary*), bibliografii i indeksu.

W przedsłowiu autorzy szkicują przyjętą definicję AI jako „zastosowanie matematyki do modelowania funkcji ludzkiego mózgu" (s. ix), przy czym lokują pytanie o ograniczenia tych modeli w kontekście koncepcji Osobliwości (Singularity). Osobliwość Techniczna jest jedną z kluczowych idei transhumanizmu, którego autorzy recenzowanej książki są przeciwnikami. Wedle tej idei rozwój techniki ma zmierzać do punktu przełomu, w którym postęp przyspieszy tak bardzo, że nie będziemy w stanie zapanować nad powstającymi artefaktami. Ten punkt jest właśnie nazywany Osobliwością. W kontekście tej książki oznacza moment powstania AI przewyższającej nas intelektualnie.

Autorzy zasadniczo rozważają tzw. silną sztuczną inteligencję (*artificial intelligence*; dalej: AGI), którą *general* definiują jako: 1. AI o poziomie inteligencji równym lub wyższym od poziomu ludzkiej inteligencji

bądź 2. AI radzącą sobie z pojawiającymi się problemami przynajmniej w porównywalnym do ludzi stopniu. Główny argument został sformułowany na wzór wnioskowania. Z przesłanek: (A1) Aby zbudować AGI, potrzebujemy rozwiązania technicznego wykazującego inteligencję co najmniej na poziomie ludzkim; (A2): Jedynym sposobem wytworzenia takiego rozwiązania jest stworzenie emulacji ludzkiego systemu neurokognitywnego przy użyciu oprogramowania; oraz wnioskowania pomocniczego: (B1) Aby stworzyć oprogramowanie emulujące działanie systemu, musimy stworzyć matematyczny model tego systemu, który pozwoli na przewidywanie jego zachowań; (B2) Nie jest możliwe zbudowanie takiego matematycznego modelu dla systemów złożonych; (B3) Ludzki system neurokognitywny jest systemem złożonym; (B4) Zatem niemożliwe jest stworzenie oprogramowania emulującego ludzki system neurokognitywny; wyciągają wniosek: (C) Wytworzenie AGI jest niemożliwe.

Rozdział 1. pt. „Wprowadzenie" przedstawia koncepcję AGI jako systemu mającego emulować oraz przewyższać (*exceed*) inteligencję ludzkiego typu, której przejawy obejmują: rozumowanie, pamięć, świadomość, uczucia i emocje, a nawet wolę i moralność (*moral thinking*). Autorzy łączą nieredukcjonistyczne podejście filozofii realistycznej ze zdroworozsądkowym spojrzeniem na świat i twierdzą, że filozofia analityczna stosuje redukcjonistyczną (tj. nieuwzględniającą złożoności ludzkiego systemu neurokognitywnego) ontologię, która nie nadaje się do opisu zagadnień kluczowych dla tej książki. Korzystają z dorobku fenomenologii realistycznej, m.in.: Husserla, Reinacha, Ingardena, Schelera i Hartmanna. Następnie wyróżniają trzy sposoby rozumienia terminu „niemożliwe": techniczny, fizyczny i matematyczny. Ten ostatni oznacza, że rozwiązanie niektórych zdefiniowanych matematycznie problemów jest niewykonalne, co wynika z apriorycznych zasad matematyki. Przytaczają argumenty innych autorów przeciwko możliwości powstania AGI i oceniają, że odzwierciedlają one stanowisko większości ekspertów od AI.

Następne 5 rozdziałów zostało wyodrębnionych jako część pt. „Własności ludzkiego umysłu". W rozdziale 2. pt. „Ludzki umysł" autorzy twierdzą, że umysł i ciało są nierozdzielne, a procesy mentalne emanują ze złożonych fizycznych procesów zachodzących w mózgu. Swój pogląd określają jako „monizm nomologiczny" (procesy mentalne i procesy fizyczne uznają za dwie nazwy dla tej samej grupy zjawisk) i porównują go z innymi stanowiskami dotyczącymi natury umysłu, m.in.: redukcyjnym i nieredukcyjnym fizykalizmem, dualizmem naturalistycznym Chalmersa, monizmem anomalnym Davidsona.

W rozdziale 3. pt. „Inteligencja ludzka i maszynowa" przytoczone zostają definicje inteligencji, wprowadzone rozróżnienie między

zdolnością (*capability*) a skłonnością (*disposition*) i między dwoma aspektami inteligencji: pierwotnym (*primal*) oraz obiektywizującym (*objectyfying*). Inteligencja pierwotna przejawia się w podejmowaniu działań będących reakcją na bodźce. Realizuje się w bieżącym zaspokajaniu podstawowych biologicznych potrzeb i nie obejmuje długoterminowych działań. Inteligencja obiektywizująca zaś wyraża się w aktach mentalnych i językowych, wiążących się z ujmowaniem samego siebie jako przedmiot tych aktów (s. 46). Dzięki inteligencji ludzie tworzą kulturę i przekształcają swe otoczenie. AI będzie musiała się więc zmierzyć ze środowiskiem powstałym w wyniku działania wielu złożonych systemów (tj. ludzi, zwierząt i środowiska naturalnego), czyli takich, których nie da się modelować matematycznie. Wobec tego niezależnie od sposobu tworzenia AI, zawsze będzie ona systemem działającym wyłącznie na podstawie zasad logiki, a więc nie będzie spełniać definicji inteligencji, która nie jest algorytmiczna – a taka jest inteligencja ludzka.

Rozdział 4. pt. „Natura ludzkiego języka" zaczyna się od stwierdzenia, że tym, co najbardziej odróżnia człowieka od zwierząt, jest zdolność mówienia i prowadzenia rozmów. Język ma być najważniejszym obserwowalnym przejawem inteligencji obiektywizującej. Rozumienie ludzkiego języka jest ważne, jeśli AI ma być użyteczna. Oprócz funkcji deskryptywnej język pełni też wiele innych funkcji ściśle powiązanych z ludzkimi uwarunkowaniami (np. wyrażanie intencji i pragnień), których nie da się w prosty sposób opisać. Duża złożoność i potencjalnie nieskończona wariantywność (*variance*) komunikacji językowej jest tu argumentem za niemożliwością stworzenia matematycznego modelu, mającego stanowić podstawę dla AGI.

Rozdział 5. pt. „Zmienność i złożoność ludzkiego języka" rozwija wątki wprowadzone w poprzedniej części. Autorzy zwracają uwagę m.in. na: intencjonalność prowadzonych konwersacji, złożoność kontekstów społecznych, strukturę jednostek językowych (słowa, gesty, emocje…) i ekonomię języka. Znaczna część interpretacji znaczeń i zamiarów odbywa się *implicite*, co nierzadko prowadzi do nieporozumień. Konwersacja oznacza interakcję między złożonymi systemami, z których każdy posiada własne intencje, zdolności i kontekst. Taka interakcja przynajmniej częściowo może zostać opisana oraz wyjaśniona, ale nie da się utworzyć matematycznego modelu, który pozwoliłby przewidywać jej dalszy przebieg.

Rozdział 6. pt. „Zachowania społeczne i etyczne" dotyczy trzech głównych zagadnień: 1. zachowania społeczne we wspólnotach, społeczeństwach i instytucjach, 2. zmiana perspektywy i intersubiektywność, 3. normy społeczne z uwzględnieniem norm prawnych i moralnych. Za

Schelerem autorzy przyjmują wartości moralne jako istniejące niezależnie od podmiotu i przez podmiot poznawalne. Oddziałują one na nasze zachowania, jednak nie w sposób deterministyczny. Normy dotyczące zachowań i społecznych kontekstów umożliwiają współpracę między jednostkami i są poznawane w procesie socjalizacji, którego nie da się zalgorytmizować. Świadomość istnienia wartości moralnych oraz uwzględnianie norm i hierarchii społecznych w swoim zachowaniu są zdolnościami występującymi tylko u ludzi.

Kolejne dwie części zostały wyodrębnione pt. „Granice matematycznych modeli". Rozdział 7. pt. „Systemy złożone" definiuje agenta AI jako: A. algorytm mający modelować dany aspekt zachowania emanującego z ludzkiego mózgu bądź B. mający modelować funkcje działające w różnych środowiskach w sposób przydatny dla ludzi. Model to przybliżone odwzorowanie wybranego wycinka rzeczywistości przy użyciu abstrakcyjnych symboli w celu: opisu (modele deskryptywne), wyjaśnienia (m. eksplanacyjne) bądź przewidywania (m. predykcyjne) aspektów tegoż wycinka. Modele opisują raczej ogólne zależności niż indywidualne przypadki. Autorzy przedstawiają klasyfikację modeli i wprowadzają rozróżnienie między symulacją (badanie, jak zmiany danych wejściowych i parametrów wpływają na dane wyjściowe) a emulacją (naśladowanie zachowania bądź funkcji jednego bytu przez inny byt). Podkreślają, że do stworzenia emulacji ludzkiej inteligencji konieczne jest stworzenie modelu obejmującego najważniejsze aspekty najbardziej skomplikowanego systemu złożonego, jakim jest continuum ciało-umysł. Autorzy przyjmują, że „continuum ciało-umysł" oznacza nierozdzielność umysłu i ciała. Struktura złożona z tych dwóch nierozłącznych części podejmuje „interakcje ze środowiskiem w celu realizacji naszych intencji, a w szczególności intencji do przetrwania i reprodukcji" (s. 308).

Rozdział 8. pt. „Matematyczne modele złożonych systemów" wskazuje ograniczenie matematyki, polegające na niemożliwości przewidywania zachowań złożonych systemów na podstawie danych dotyczących ich struktury i zachowania w przeszłości ze względu na przypadkowość zmian zachodzących w obrębie tejże struktury (pojawianie się i zanikanie elementów) jak również losowość zmian interakcji i wzajemnego powiązania między jej elementami. Przewidywanie na podstawie opisu obecnego stanu również jest niemożliwe ze względu na ogromną liczbę elementów oraz dużą zależność od otoczenia. Autorzy omawiają różne sposoby budowania modeli na potrzeby AI (m.in. równania różniczkowe, modele stochastyczne, modele regresyjne) i wskazują, że niezależnie od ich użyteczności w wąskich dziedzinach, żaden z nich nie może być podstawą do utworzenia AGI.

Ostatnie pięć rozdziałów zostało wydzielone pod wspólnym tytułem „Ograniczenia i potencjał AI". Rozdział 9. pt. „Dlaczego nie będzie inteligencji maszynowej" podważa filozoficzne argumenty za możliwością powstania AGI. Autorzy wykazują błędne założenia Chalmersa argumentu emulacji, błąd rozumowania w argumencie z ewolucji (odwołanie się do równoważności, która nie zachodzi) i nieprawidłowe zdefiniowanie inteligencji w argumencie, że Osobliwość może zostać osiągnięta bez AGI. Zwracają uwagę, że maszyny nie mogą posiadać popędów (*driveness*), intencji ani woli, a co za tym idzie nie mogą być świadome (nawet jako zombie). Maszyny nie są w stanie prowadzić teoretycznych rozumowań ani rozwinąć inteligencji obiektywizującej.

Rozdział 10. pt. „Dlaczego maszyny nie opanują ludzkiego języka" rozpoczyna się od wskazania, że opanowanie (*mastery*) języka jest warunkiem koniecznym i wystarczającym dla zaistnienia AGI. Język jest dowodem zaawansowania ludzkiej inteligencji. Autorzy przedstawiają listę kryteriów dotyczących zdolności językowych potencjalnie przejawianych przez maszyny, których spełnienie byłoby mocnym argumentem za tym, że mamy do czynienia z AGI. Język obejmuje nie tylko mowę, ale także komunikację niewerbalną, która wymaga rozumienia bodźców wizualnych i sensorycznych. Niemożliwe jest utworzenie matematycznego modelu ludzkiego języka, jako że jest on złożonym systemem, będącym emanacją innego złożonego systemu (tj. continuum ciało-umysł). Znaczenia konkretnych fraz zależą od danych kontekstów, które także są złożone.

Rozdział 11. pt. „Dlaczego maszyny nie opanują interakcji społecznych" podsumowuje postawione w poprzednich częściach tezy dotyczące niemożliwości utworzenia modelu ludzkiego języka ani modelu interakcji społecznych. Maszyny nie posiadają umysłów i nie są w stanie wykazać rozumienia sytuacji z punktu widzenia innych podmiotów. Z tego względu ich działania nie mogą być oceniane jako podlegające prawu ani jako to prawo egzekwujące. Działania moralne muszą uwzględniać złożony system wartości, którego nie da się emulować. Każdy człowiek na swój sposób dokonuje sądów moralnych, nie da się zatem utworzyć ogólnego modelu ludzkiej moralności. Zatem maszyny nie mogą „rozumieć" moralności.

Rozdział 12. pt. „Cyfrowa nieśmiertelność" jest kulminacją argumentacji przeciwko możliwości powstania Osobliwości. Przywołane zostały tezy dotyczące oddzielenia umysłu od ciała w celu przeniesienia „człowieka" do cyfrowego świata jak również nawoływania Bostroma aby filozofowie i matematycy porzucili swoje dotychczasowe zajęcia (bo AI wykona je lepiej) a zamiast tego skupili się na pracy na rzecz

przeciwdziałania niszczycielskim skutkom Osobliwości. Autorzy argumentują za: niemożliwością emulowania systemów złożonych przez maszyny Turinga, nierozdzielnością umysłu i ciała, niewykonalnością Emulacji Całego Mózgu (*Whole Brain Emulation*), niemożliwością kognitywnego ulepszania (*enhancement*) człowieka oraz nonsensownością tez o „eksplozji inteligencji", powstania nikczemnej AI i przedłużaniu życia przez wykorzystanie nanobotów i inżynierii genetycznej. Na stronie 286 pojawia się argument nawiązujący do głównego argumentu tego tekstu (zawartego w „Przedsłowiu") a odnoszący się do cyfrowej nieśmiertelności. Z przesłanek: (A) Cyfrowa nieśmiertelność wymaga komputerowej symulacji indywidualnego ludzkiego mózgu; (B) Ludzkie umysły są systemami złożonymi; (C) Komputerowa symulacja wymaga odpowiedniego matematycznego modelu symulowanego systemu; (D) Niemożliwe jest stworzenie adekwatnych (tj. nieuproszczonych) matematycznych modeli złożonych systemów; wyciągają wniosek: cyfrowa nieśmiertelność jest niemożliwa. Na końcu autorzy parafrazują apel Bostroma, aby filozofowie, matematycy oraz „wszyscy zdrowi na ciele i umyśle" zajmowali się tematami jak najbardziej odległymi od „science fiction à la Bostrom" (s. 287).

Rozdział 13. pt. „Wiosna AI jest wieczna" zaczyna się od twierdzenia, że aby uniknąć frustracji związanych z nadejściem kolejnych „zim AI" oraz odnosić korzyści z badań nad AI należy się skupić na „wąskiej AI", zajmującej się ściśle określonymi problemami, które da się rozwiązać. Autorzy podkreślają, że jest ona tylko narzędziem, którego użyteczność jest ogromna, choć ograniczona do pewnego typu zadań. W każdym sektorze gospodarki obecne są obszary niebędące systemami złożonymi, które mogą zostać skutecznie zautomatyzowane przy pomocy AI. Dzięki temu poprawiona zostanie efektywność – do wykonywania pewnych zadań potrzeba będzie mniej ludzi, ale nie uda się całkowicie wyeliminować ludzkiej pracy. Autorzy zaznaczają, że największym wyzwaniem związanym z rozwojem AI nie jest zastąpienie ludzkiej inteligencji ani zagrożenia ze strony Osobliwości, lecz pomoc w znalezieniu zajęcia ludziom, którzy zostaną zastąpieni przez algorytmy. Rozwój AI przebiega tak samo jak rozwój każdego innego wytworu techniki.

Tekst jest napisany jasnym językiem, a wyjaśnienia omawianych przez autorów poglądów znacząco ułatwiają lekturę. Natomiast w rozdziałach jest wiele odniesień do innych części, np. na s. 26 (rozdział 2.2.2) znajduje się odniesienie do rozdziału 7, a na s. 29 do rozdziałów 8 i 12. Świadczy to o ciągłości wywodu i przemyślanej, choć nieco nieoczywistej strukturze. Pierwsza część (strony 21-107) zawiera obszerny kontekst i stanowi wprowadzenie do wątków poruszanych w kolejnych dwóch częściach

tekstu. Stąd np. wątki poruszane w rozdziale 3. są rozwijane w r. 9., wątki z r. 4. i 5. w r. 10., a wątki z r. 6. w r. 11. Sprawia to, że wprowadzenie do danego problemu i jego rozwinięcie w kontekście AI dzieli czasem kilkadziesiąt stron. Znacząco zwiększa to trudność czytania książki ze względu na konieczność „przeskakiwania" do odpowiednich fragmentów. Z drugiej strony (taki był zamysł autorów) osoby obeznane z tematem mogą rozpocząć lekturę od drugiej części tekstu, aby zapoznać się z głównymi argumentami autorów bądź od trzeciej części, aby poznać ich poglądy nt. AGI. Jednak w tym przypadku lekturę mogą utrudniać odnośniki do wcześniejszych części tekstu. Podrozdział 12.3.4 nosi ten sam tytuł co rozdział 12., co naraża autorów na zarzut, że poprzednie podrozdziały są zbędne.

Landgrebe i Smith odwołują się do fizyki i procesów fizycznych jako podłoża procesów mentalnych. Przyjęli dwa założenia, których nie wyrażają wprost: 1. Tylko jedno stanowisko wobec relacji ciało-mózg może być prawdziwe; 2. Prawdziwe jest stanowisko przyjęte przez nich (continuum ciało-umysł). Powołują się na wiele naukowych teorii. Sprawia to, że miejscami tekst bardzo zagłębia się w szczegóły, przez co bywa trudniejszy w odbiorze. Ilustruje to rozdział 6, poruszający wiele kwestii związanych z komunikacją oraz normami społecznymi i moralnymi. Świadczy to natomiast o obszernej wiedzy autorów oraz wysokim poziomie merytorycznym tekstu. Główną część argumentacji można streścić następująco: AGI nie może powstać, bo do jej utworzenia potrzebny jest matematyczny opis (model), a matematyka nie jest w stanie nawet w uproszczony sposób uchwycić tak złożonych i losowo zmieniających się struktur jak ludzkie mózgi. Innymi słowy: opis nigdy nie jest w stanie wyczerpać rzeczywistości, a zatem na jego podstawie nie można utworzyć wiernej repliki realnej rzeczy. Wydaje się, że co do tego panuje powszechna zgoda, przynajmniej na gruncie filozofii. Złośliwy dysputant mógłby jednak stwierdzić, że autorzy wykazują co najwyżej niemożliwość powstania AGI na gruncie przyjętej przez siebie ontologii, a zatem dyskutują sami ze sobą, ponieważ nie podważają założeń transhumanizmu. Ten zaś skupia się raczej na opisie struktury informacyjnej mózgu (tzw. wzorcu informacyjnym) niż na matematycznym modelowaniu jego funkcji. Szkoda, że autorzy nie poszerzyli swoich analiz tej kwestii. Transhumanizm nie popiera konkretnego stanowiska w kwestii metafizyki. Na podstawie różnych tekstów dopuszczalna jest interpretacja, iż jest to dualizm (materialne ciało połączone z niematerialną informacją), hylemorfizm (niematerialna informacja kształtująca materialne ciało) a nawet idealizm (istotna jest tylko informacja, materialna forma nie ma znaczenia, a może nawet jest niepotrzebna).

Autorzy przedstawili tylko ogólną koncepcję AGI, co jest słabością tego tekstu. Określenie „emulowania i przekraczania ludzkiej inteligencji" jest zbyt ogólne. Powoływanie się na badania głównego nurtu nad AGI (*mainstream AGI research*; s. 57) jest nieco na wyrost, bowiem nie istnieje ogólnie przyjęta definicja AGI. Autorzy dyskutują nie tyle przeciw możliwości powstania AGI jako takiej, ile przeciw możliwości wytworzenia „syntetycznego umysłu" np. w postaci Osobliwości. *Clou* argumentacji odnosi się do (nie)możliwości stworzenia wirtualnego człowieka, co jest postulowane przez niektórych transhumanistów jako „cyfrowa nieśmiertelność". Jednak, zgodnie z Kurzweila koncepcją tożsamości narracyjnej, wystarczy, że AI będzie w stanie odpowiednio naśladować zachowania osoby, którą odwzorowuje. Nie wymaga to dokładnych predykcji, jakie postulują autorzy, ponieważ jest to model deskryptywny – w odpowiednim przybliżeniu opisuje działania danej osoby, ale nie może ich przewidywać. Zatem w przypadku osiągnięcia „nieśmiertelności" argumentacja autorów jest nietrafiona. Natomiast kwestia Osobliwości jest otwarta na dyskusje. Jednak wszelkie dysputy muszą się rozpocząć od podania definicji, których tutaj zabrakło.

Wydaje się, że emulacja ludzkiego mózgu na potrzeby przemysłu i nauki nie musi być aż tak szczegółowa, jak chcą tego Landgrebe i Smith. Moim zdaniem modelowanie na poziomie pojedynczych cząstek i komórek jest niepotrzebne – stanowi raczej figurę retoryczną (hiperbolę) niż faktyczny wymóg. AI dzięki swojej specjalizacji w wąskich dziedzinach (np. modelowanie białek – AlphaFold, inżynieria materiałowa – GNoME) już teraz dostarcza użytecznych rozwiązań. Ponadto jej rezultaty są lepsze niż te uzyskane przez najlepsze zespoły badawcze. Z argumentacji autorów wynika, że matematycznie niemożliwe jest sformułowanie modelu opisującego złożony system, jednak nie uchyla to argumentu za możliwością stworzenia takiego systemu. GPT-2 został zbudowany na podstawie modelu obejmującego 1,5 mld parametrów, GPT-3 miał ich już 175 mld, GPT-4 ma ich 1,8 bln a ponadto (dzięki połączeniu z modelem DALL-E 3) oprócz tekstu potrafi interpretować i generować grafikę. Kolejne modele będą jeszcze większe i będą dysponować jeszcze większymi możliwościami. Paradygmat tworzenia zaawansowanych AI w przyszłości może ulec zmianie, np. AGI może powstać jako sieć integrująca wiele wąsko wyspecjalizowanych AI. Byłby to rodzaj złożonego systemu, który w teorii mógłby wykonywać większość zadań lepiej niż najlepsi specjaliści, a którego nie można by matematycznie opisać, bo sam będzie zbyt złożony. Zatem twierdzenie, że AGI nie może nigdy powstać, ponieważ nie jesteśmy w stanie matematycznie opisać naszych mózgów, jest w najlepszym razie przedwczesne, zwłaszcza jeśli przyjmiemy tezę

głoszoną przez zwolenników nadejścia Osobliwości, że świadomość i procesy kognitywne mogą się pojawić samoistnie w odpowiednio skomplikowanych systemach. Warto także podkreślić, że argumentacja autorów odnosi się do realizacji AGI na „uniwersalnej maszynie Turinga", której uproszczonym przykładem są współczesne komputery cyfrowe. Jednak (przynajmniej w teorii) możliwe są inne podejścia. Podnoszone w książce argumenty nie obejmują m.in. komputerów analogowych, sprzętowych sieci neuronowych, ani hipotetycznych konstrukcji pokroju komputerów fotonicznych, które nie są maszynami Turinga. Ponadto znane są tezy, że do powstania AGI konieczne jest oddziaływanie ze środowiskiem. Taki rodzaj „ucieleśnionej AI" również nie musi działać jak maszyna Turinga.

Autorzy polemizują m.in. z koncepcją, wedle której sieci neuronowe naśladują strukturę i funkcje biologicznych mózgów. Jednak nie zauważają, że wbrew temu, co twierdzą komputacjoniści i funkcjonaliści, maszyny nie naśladują działań ludzkiego umysłu lecz rezultaty tych działań. Zatem „prawdziwa AI" niekoniecznie musi być skonstruowana tak, jak nasze mózgi. Myślę, że nie musi nawet posiadać dokładnego rozumienia języka – nierzadko my sami mamy problem z odczytaniem intencji innych ludzi. Na co dzień spotykamy się z sytuacjami, gdy nasi rozmówcy (lub my sami) muszą wyjaśniać, o co im chodziło. Niewykluczone, że lepsza od nas AGI również będzie miała takie problemy. Może nawet się okazać, że mimo większej inteligencji i rozeznania złożonych problemów w ogóle nie będzie mogła się z nami komunikować. Przydatne byłoby ujęcie innego rodzaju niemożliwości niż wskazane w książce trzy definicje (s. 8), a mianowicie: niemożliwe jest takie zrozumienie ludzkiego języka i norm społecznych, które pozwoliłoby na bezbłędną komunikację.

Warto byłoby, aby autorzy rozważyli metafizyczne konsekwencje przyjętego przez siebie stanowiska (monizm nomologiczny), zwłaszcza w kwestii determinizmu. Wedle tego, co nakreślili, wszelkie kognitywne funkcje ludzkiego umysłu emanują z układów cząstek, a sama emanacja jest procesem fizycznym (s. 30). Opowiadają się za niedeterministycznym charakterem ludzkich działań, np. wyrażając niemożliwość predykcji przebiegu interakcji między ludźmi w postaci rozmowy (s. 89). Z drugiej strony twierdzą, że AGI dla osiągnięcia poziomu ludzkiego lub ponadludzkiego musiałaby być tak skonstruowana, że jej działanie byłoby przewidywalne przy użyciu narzędzi logiki oraz praw fizyki. Trudno stwierdzić czy według autorów ludzie działają w sposób deterministyczny oraz (co się z tym wiąże) czy posiadają wolną wolę – skoro jesteśmy tylko układami cząstek, których wszystkie elementy da się opisać prawami fizyki, to na jakiej podstawie zachodzi autodeterminacja do działania? Skoro

wszystkie procesy kognitywne są wynikiem oddziaływań między cząstkami tworzącymi mózgi (continuum ciało-mózg), to czy „moje" działania faktycznie są moje? Jak istnieją wartości (*values*), skoro są niezależne od podmiotu, a jednocześnie wyznaczają zasady naszego zachowania (s. 103)? Te kwestie nie mają większego wpływu na sedno argumentacji autorów, mogą obniżać jej siłę, a na pewno pozostawiają pewien niedosyt.

W wielu miejscach książka sprawia wrażenie zlepku wątków filozoficznych i naukowych (a dokładniej: neuronauk). Prowokuje to do pytania, czy przenoszenie filozoficznych sporów na grunt nauk przyrodniczych jest uprawnione pod względem metodologicznym i jaki jest status poznawczy otrzymanych wyników. Jest to ciekawy temat na odrębne studium. Natomiast z pewnością książka jest interesującą propozycją dla wszystkich osób zainteresowanych filozofią umysłu, transhumanizmem, cyfrową nieśmiertelnością oraz możliwością powstania silnej sztucznej inteligencji (AGI). Warto po nią sięgnąć przede wszystkim ze względu na szeroki kontekst dyskusji podjętej przez autorów. Obszerna bibliografia zawiera odwołania do ważnych prac dotyczących umysłu, inteligencji, komunikacji, zachowań społecznych, modeli matematycznych i wielu innych. Mniej zaawansowani czytelnicy będą mogli poszerzyć swoją wiedzę, a ci bardziej zaawansowani będą mieli okazję do jej uzupełnienia.