2024 › nr 1 (55)

filozofuj!

magazyn popularyzujący filozofię

Cena 15 zł
(w tym 8% VAT)
Nakład 2500 egz.

SZTUCZNA INTELIGENCJA

ISSN 2392-2249
Indeks nr 416851

filozofuj.eu
redakcja@filozofuj.eu

Wydawnictwo Academicon

Drodzy Czytelnicy,

jedną z największych korzyści, jaką można odnieść z uprawiania filozofii, jest nawet nie tyle uzyskanie ostatecznych odpowiedzi na nurtujące nas pytania, ile sama elastyczność myślenia, która pozwala nam w nieszablonowy sposób spojrzeć na nowe problemy, które stają na naszej intelektualnej drodze. A przecież niekiedy mamy wrażenie, że szalone tempo zmian, jakie zachodzą w świecie, nie daje nam żadnych szans. Warunki funkcjonowania w społeczeństwie, kulturze, mediach potrafią być zrewidowane w relatywnie krótkim czasie i jeśli jakakolwiek dyscyplina pozwala mieć nadzieję na zorientowanie się co do skali i charakteru owych zmian, to z pewnością jest to filozofia.

Oczywiście trzeba być bardzo ostrożnym, kiedy ogłasza się, że jakiś nowy wynalazek lub osiągnięcie w dużym stopniu zmienia reguły gry. Jednak uznanie, że możliwości sztucznej inteligencji stanowią okoliczność, na którą powinniśmy intelektualnie zareagować, nie wydaje się obarczone wielkim ryzykiem. Ze względu na aktualność tego tematu zajmujemy się nim w najnowszym numerze pisma. Nie gwarantujemy, że uśmierzymy Wasz lęk przed sztuczną inteligencją albo że wzbudzimy w Was nadzieję na osiągnięcia, które może zdobyć ludzkość, używając sztucznej inteligencji. Możemy zaproponować Wam jedynie prawdę, a więc rzetelną debatę na temat szans i zagrożeń, jakie wiążą się z rozwojem sztucznej inteligencji (SI, AI).

Stawiając pytania o to, czy sztuczna inteligencja może być faktycznie podobna do typowych podmiotów poznających, czy typowo ludzka inteligencja znajduje się na z góry przegranej pozycji i będzie tracić na znaczeniu, w racjonalny sposób wyrażamy wątpliwości, które rodzą się wraz z kolejnymi odkryciami na polu sztucznej inteligencji. Jeśli rację ma Stanisław Lem, gdy w jednym ze swoich opowiadań przewiduje, że człowiek ze względu na swoją niedoskonałość jest w stanie wygrać pojedynek z maszyną, to być może nie mamy powodu faktycznie obawiać się sztucznej inteligencji. Jednak nawet wtedy musimy zrobić wszystko, żeby spróbować ją zrozumieć. Mamy nadzieję, że pomogą w tym teksty z naszego numeru: o tym, jak działa sztuczna inteligencja (Karol Draszawka); o tym, czy jest możliwa świadoma SI (Robert Poczobut); o obawach (Piotr Kulicki i Barry Smith), perspektywach (*Sztuczna inteligencja – zagrożenie czy szansa dla demokracji?*) i wyzwaniach (*Edukacja w dobie sztucznej inteligencji*) związanych ze sztuczną inteligencją oraz o etyce SI (Artur Szutta). Obrazu problemu dopełni zaś niezwykle ciekawa rozmowa z Lucianem Floridim. Kto jest spostrzegawczy, dostrzeże, że duży wkład w powstanie numeru wniosła sama SI – zilustrowała wszystkie artykuły i stworzyła jeden z tekstów tematycznych.

Redakcja

filozofuj!
2024 › nr 1 (55)
magazyn popularyzujący filozofię

STYCZEŃ

■ **3 stycznia 1904 r.** – we Lwowie przyszła na świat **Izydora Dąmbska**, polska filozofka, logiczka, tłumaczka i epistemolożka, przedstawicielka Szkoły Lwowsko-Warszawskiej. Dąmbska specjalizowała się w historii filozofii, semiotyce, metodologii nauk oraz teorii poznania. W obszarze semiotyki zajmowała się m.in. problematyką nazw pustych, imion własnych oraz zdań warunkowych. Jej badania obejmowały też semiotyczne funkcje milczenia, a także opracowanie kryteriów rozumienia. W metodologii poszukiwała m.in. granicy między nauką a światopoglądem, interesowała się istotą rozumowania przez analogię i sformułowała nowoczesną koncepcję praw nauki. Do jej najważniejszych publikacji należą *Dwa studia z teorii naukowego poznania* (1962) oraz *Wprowadzenie do starożytnej semiotyki* greckiej (1980). (M. B.)

Izydora Dąmbska

■ **5 stycznia 2001 r.** – zmarła brytyjska filozofka, **Elisabeth Anscombe**. Była jedną z najwierniejszych uczennic Ludwiga Wittgensteina, którego pierwszy raz spotkała w 1942 r. w Cambridge. Miała też okazję usłyszeć jego słynne ostatnie słowa: „Eliza, ja zawsze kochałem prawdę". Po śmierci mistrza przetłumaczyła *Dociekania filozoficzne* na język angielski oraz zajęła jego katedrę w Cambridge. Interesowała się problemami z niemal każdej dziedziny filozofii, jednakże największy wkład wniosła w badania z zakresu teorii działania (słynna monografia *Intention* z 1957 r.) (A. M.)

Elisabeth Anscombe

■ **20 stycznia 1922 r.** – urodził się **John Hick**, brytyjski filozof religii i teolog. Bronił on hipotezy pluralizmu religijnego, zgodnie z którą wielkie religie świata są różnymi sposobami doświadczania Ostatecznej Rzeczywistości. W swojej koncepcji Hick wykorzystał Kantowskie rozróżnienie na fenomeny (zjawiska) i noumeny (rzeczy same w sobie). Dzięki religiom doświadczamy jedynie rzeczywistości fenomenalnej, natomiast Rzeczywistość sama w sobie jest niepoznawalna i niewyrażalna. (T. K.)

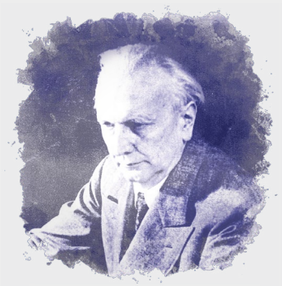
LUTY

■ **8 lutego 412 r.** (według jednej z wersji) – urodził się **Proklos**, scholarcha Akademii Platońskiej, przeciwnik chrześcijaństwa, uważany za jednego z ostatnich wielkich greckich filozofów. Odegrał istotną rolę w rozpowszechnieniu się idei neoplatońskich w Bizancjum, a później świecie islamu. Uważał, że do prawdy prowadzi nie tylko rozum, ale również właściwie zrozumiany mit, a także wiara stanowiąca zjednoczenie człowieka oraz wszystkich bogów z niepoznawalnym i najdoskonalszym Absolutem. To właśnie mistyczne zjednoczenie uznawał za naczelny cel życia. (R. W.)

■ **8 lutego 1999 r.** – zmarła szkocka filozofka i pisarka, **Iris Murdoch**; platonistka i egzystencjalistka oraz jedna z niewielu przedstawicielek brytyjskiej filozofii dystansujących się od tradycji analitycznej. Jej zainteresowania filozoficzne ogniskowały się wokół problematyki religii i moralności. Głosiła potrzebę stworzenia nowego słownika doświadczenia moralnego; jej poglądy były bliskie przekonaniu Simone Weil, że moralność jest sprawą sposobu patrzenia, a nie woli. (A. M.)

■ **13 lutego 1869 r.** – **Fryderyk Nietzsche** rozpoczął pracę na Uniwersytecie w Bazylei w Szwajcarii. Już rok później, w 1870 r., mimo młodego wieku – miał wówczas zaledwie 25 lat – został awansowany z pozycji profesora nadzwyczajnego do profesora zwyczajnego filologii klasycznej. W tym czasie nie zaczął jeszcze nawet pisać swojej rozprawy doktorskiej. Stanowisko zawdzięczał profesorowi Friedrichowi Wilhelmowi Ritschlowi, który tak bardzo docenił jego esej dotyczący greckiego poety, Teognisa, że opublikował go w swoim czasopiśmie naukowym „Rheinisches Museum" i promował Nietzschego jako fenomen filologii klasycznej. (M. B.)

■ **23 lutego 1883 r.** – w Oldenburgu urodził się niemiecki filozof i psychiatra **Karl Jaspers**. Był profesorem filozofii w Heidelbergu i Bazylei. Uważany jest za jednego z twórców egzystencjalizmu teistycznego. Ważnym pojęciem jego koncepcji filozoficznej jest pojęcie sytuacji granicznych, do których należą wina, walka, cierpienie i śmierć. Mają one źródło w antynomicznej strukturze świata. Gdy człowiek znajdzie się w takiej sytuacji, uświadamia sobie, że nie jest tak potężny, jak mu się wydawało, lecz kruchy, bezradny i słaby. Do dzieł Jaspersa należą: *Filozofia egzystencji*, *Wprowadzenie do filozofii: dwanaście odczytów radiowych*, *Rozum i egzystencja. Nietzsche a chrześcijaństwo* czy *Autorytety: Sokrates, Budda, Konfucjusz, Jezus*. (L. G.)



Karl Jaspers

Opracowanie:
M. B. – Milena Bartoszewska,
L. G. – Liliana Gołąb
T. K. – Tomasz Kaliński,
A. M. – Aleksandra Majdak,
R. W. – Rafał Wąż

6 Sztuczna inteligencja, czyli co? > *Karol Draszawka*
Według jednej z definicji sztucznej inteligencji (SI) jest ona „częścią informatyki zajmującą się projektowaniem inteligentnych systemów komputerowych, [...] które wykazują cechy kojarzone z inteligencją w ludzkim zachowaniu – rozumieniem języka, uczeniem się, rozumowaniem, rozwiązywaniem problemów i tak dalej" (Barr, Feigenbaum 1981). W tym eseju, pisząc „SI", będę miał na myśli inteligentne systemy komputerowe, a nie dziedzinę informatyki zajmującą się nimi. Na czym polega działanie owych systemów?

9 Czy sztuczna świadomość jest możliwa? > *Robert Poczobut*
W filozofii sztucznej inteligencji stawiamy pytanie o możliwość istnienia umysłów niebiologicznych, zbudowanych na bazie innego substratu (budulca) niż struktury węglowo-białkowe. Chcielibyśmy wiedzieć, czy świadomość u swych podstaw musi być organiczna, czy też jej istnienie jest w jakimś sensie niezależne od substratu.

12 Sztuczna inteligencja jako podróż w Nieznane > *Piotr Kulicki*
O sztucznej inteligencji i jej spektakularnych postępach napisano w ostatnich latach bardzo wiele, także w „Filozofuj!", i właściwie trudno coś nowego tu dodać. Dlatego chciałbym zwrócić uwagę raczej na to, jakie są reakcje ludzi na jej postępujący rozwój. Myślę, że jest to część szerszego zjawiska – reakcji ludzi na to, co nieznane, niezrozumiałe czy nowe. Reakcje te od wieków odnotować można w odbiorze świata i w zasadzie od wieków się nie zmieniają.

15 Dlaczego ChatGPT nigdy nie będzie rządzić światem > *Barry Smith*
Jobst Landgrebe i ja w naszej najnowszej książce *Why Machines Will Never Rule the World* (*Czemu maszyny nigdy nie będą rządzić światem*) argumentujemy, że wysiłki wielu osób ze społeczności sztucznej inteligencji zmierzające do stworzenia ogólnej sztucznej inteligencji (*artificial general intelligence*, AGI) są skazane na niepowodzenie. Mówiąc w tym przypadku o AGI, mamy na myśli maszynę, która wykazywałaby zdolności poznawcze równoważne lub nawet przewyższające ludzkie.

17 Czym zajmuje się etyka sztucznej inteligencji? > *Artur Szutta*
Etyka sztucznej inteligencji (AI) to dziedzina etyki szczegółowej. Jej zakres i treść wyznaczają pytania, które stawiają etycy, badając zagadnienie sztucznej inteligencji. Oto główne kwestie oraz idee, jakimi zajmują się etycy AI.

20 Edukacja w dobie sztucznej inteligencji > *Krzysztof Saja*
System edukacji w Polsce powinien ulec radykalnym i głębokim zmianom. Opinia ta, podzielana zapewne przez bardzo wiele młodych osób, wydaje się szczególnie trafna w kontekście wykładniczego rozwoju nowych technologii, czwartej rewolucji przemysłowej oraz sztucznej inteligencji. Jak powinna wyglądać szkoła wolnych ludzi, dla których prace wykonywać będą inteligentne maszyny?

Ilustracja na okładce: © by Mira Zyśko

22 Sztuczna inteligencja – zagrożenie czy szansa dla demokracji? > *ChatGPT*
Czy sztuczna inteligencja (SI) będzie cybernetycznym wsparciem dla demokracji, czy też raczej okaże się jej cyfrowym wrogiem? Rozważmy powszechne obawy związane z SI, jak i możliwe szanse, jakie ona daje.

24 Nadal potrzebujemy okrzyku „Eureka!" > *wywiad z Lucianem Floridim, jednym z największych autorytetów we współczesnej filozofii, twórcą filozofii informacji i jednym z głównych interpretatorów rewolucji cyfrowej.*

Narzędzia filozofa
28 Eksperyment myślowy: E-Daimonion > *Artur Szutta*
30 Kurs logiki: #18. Presupozycje – ich wykrywanie i właściwości > *Krzysztof A. Wieczorek*

Filozofia nauki
32 Krótka historia atomu: Dialog 4. Kinetyczno-molekularna teoria materii > *Andrzej Łukasik*

Filozofia społeczna
34 Śniadanie kontynentalne: #18. Sztuczna inteligencja na wolności? > *Tomasz Kubalica*

W 12 odcinkach
36 Kurs ontologii: #5. Własności i zbiory > *Arkadiusz Chrudzimski*

Filozofia w filmie
39 *2001: Odyseja kosmiczna* > *Piotr Lipski*
40 Fragment z klasyka: Spinaczowa Sztuczna Inteligencja

Filozofia w literaturze
41 Algorytm a mądrość praktyczna > *Natasza Szutta*
42 Fragment z klasyka: Zapasowa kopia

Felieton
44 Kopernik, Darwin, Turing > *Jan Woleński*
46 Wymóżdżać się sztuczną inteligencją > *Adam Grobler*
47 Robot na plaży, czyli nowy test Turinga > *Jacek Jaśtal*

Satyra
50 List do Jej Magnificencji AI > *Piotr Bartula*

Filozofia w szkole
52 Roboty w służbie ludzi > *Dorota Monkiewicz*

Z półki filozofa...
53 Ciekawość, czyli pierwszy stopień do wiedzy > *Piotr Biłgorajski*
Pochwała przyjaźni > *Paweł Sikora*

54 Filozofia z przymrużeniem oka

Szanowna Czytelniczko,

czasopismo „Filozofuj!” powstaje wysiłkiem osób, którym leży na sercu popularyzacja filozofii. Chcemy, aby było ono **dostępne bezpłatnie online** i dzięki temu mogło docierać do jak najszerszego kręgu czytelników. Jego przygotowywanie rodzi jednak niemałe koszty (skład i korekty, projektowanie grafik, utrzymanie strony czasopisma). Twoje wsparcie pozwoliłoby nam rozwijać czasopismo.

Z góry

Wsparcie / Crowdfunding

Fundacja Academicon

Szanowny Czytelniku,

Jeśli chcesz wesprzeć tę inicjatywę dowolną kwotą (1 zł, 2 zł lub inną), kliknij poniższy przycisk przekierowujący na naszą stronę filozofuj.eu/wsparcie:

Chcę wesprzeć „Filozofuj!”

dziękujemy!

filozofuj.eu redakcja@filozofuj.eu

Sztuczna inteligencja, czyli co?

Słowa kluczowe: sztuczna inteligencja, uczenie maszynowe, sieci neuronowe, nietransparentność

Karol Draszawka

Pracuje na Wydziale Elektroniki, Telekomunikacji i Informatyki Politechniki Gdańskiej, gdzie opowiada nieco dokładniej o SI. Opiekuje się Kołem Naukowym „Gradient", skupiającym pasjonatów uczenia maszynowego. Domuje z żoną i dwójką uroczych szkrabów.

Według jednej z definicji sztucznej inteligencji (SI) jest ona „częścią informatyki zajmującą się projektowaniem inteligentnych systemów komputerowych, […] które wykazują cechy kojarzone z inteligencją w ludzkim zachowaniu – rozumieniem języka, uczeniem się, rozumowaniem, rozwiązywaniem problemów i tak dalej" (Barr, Feigenbaum 1981). W tym eseju, pisząc „SI", będę miał na myśli inteligentne systemy komputerowe, a nie dziedzinę informatyki zajmującą się nimi. Na czym polega działanie owych systemów?

SIECI BAYESOWSKIE (PRZEKONANIOWE) – narzędzie służące wnioskowaniu w sytuacji niepewności, stanowiące zwięzłą reprezentację rozkładów prawdopodobieństwa wielu zmiennych; nazwa od nazwiska Thomasa Bayesa (ur. 1702, zm. 1761), brytyjskiego matematyka, którego uważa się za jednego z twórców wzoru na prawdopodobieństwo warunkowe.

Termin „inteligentny" z powyższej definicji jest rozumiany szeroko: wystarczy, żeby systemy *wykazywały* pewne cechy, które kojarzymy z ludzką inteligencją, aby określić je mianem SI. Jeśli np. program komputerowy gra w szachy tak, „jakby rozumiał" grę, i trudno go pokonać, to to wystarczy, by mówić o nim jako o sztucznej inteligencji. Właściwie wystarczy, aby dany system/moduł wykazywał choćby jedną z cech inteligencji ludzkiej, aby mówić o nim jako o SI – program szachowy nie musi rozpoznawać znaków drogowych, a system do rozpoznawania znaków drogowych nie musi znać gramatyki języka polskiego czy angielskiego. Z powyższego powodu tego typu systemy (*de facto* wszystkie obecnie istniejące systemy SI) nazywa się specjalizowaną lub wąską SI, w odróżnieniu od (wciąż jeszcze hipotetycznej) ogólnej sztucznej inteligencji (ang. *Artificial General Intelligence*, także *human-level AI*), która radziłaby sobie z każdym intelektualnym zadaniem, z jakim radzi sobie przeciętny człowiek.

Rodzaje SI

Można mówić o kilku rodzajach SI. Pierwszy z nich obejmuje systemy zdolne do wnioskowań logicznych. Istnieją dwa podrodzaje takich systemów: 1) wnioskujące w warunkach wiedzy pewnej, wyciągające wnioski z wprowadzonego zbioru faktów (wyrażonych symbolicznie) i reguł wnioskowania, oraz 2) wnioskujące w warunkach wiedzy niepewnej (opierające się na np. tzw. sieciach bayesowskich[1]), oszacowujące prawdopodobieństwa zdarzeń.

Do SI zaliczamy także algorytmy przeszukiwania grafów, używane np. do wyznaczania najkrótszej drogi między wybranymi dwoma punktami na mapie czy do znajdowania ruchu w grach planszowych tak, by zminimalizować własne straty przy założeniu maksymalnie dobrego ruchu przeciwnika. Ciekawym obszarem SI są badania nad metodami optymalizacji, np. algorytmami ewolucyjnymi, służącymi do odnajdywania satysfakcjonujących rozwiązań nietrywialnych problemów – w przemyśle czy logistyce.

Kolejna grupa algorytmów odnosi się do tzw. uczenia maszynowego, które – szczególnie uczenie głębokie (ang. *deep learning*), bazujące na tzw. sztucznych sieciach neuronowych – cieszy się obecnie największym zainteresowaniem. Jemu przyjrzymy się nieco bliżej.

Ilustracja: ChatGPT

Uczenie maszynowe

Co to znaczy, że jakiś program „się uczy"? Co to są sztuczne sieci neuronowe? Jak głębokie jest „uczenie głębokie"? Próbując odpowiedzieć na te pytania, skupmy się na najważniejszym sposobie uczenia maszynowego, na tzw. uczeniu nadzorowanym (ang. *supervised learning*). Stanowi ono podstawę zdecydowanej większości najbardziej przydatnych systemów wyposażonych w SI np. wspierających rozpoznawanie mowy, rozpoznawanie typów obiektów na zdjęciach, systemów wspomagających diagnostykę medyczną itd.

Wspólnym mianownikiem takich systemów jest klasyfikacja, przyporządkowanie właściwej klasy danemu obiektowi. Lista możliwych klas obiektów jest z góry ustalona i niezmienna. Np. krótkiej próbce dźwięku (rzędu milisekund) przypisuje się jeden z kilkudziesięciu tzw. fonemów występujących w języku mówionym, fragmentowi zdjęcia przypisuje się klasę obiektu na nim występującego (z narzuconej listy rozpoznawanych klas), wynikom laboratoryjnym lub obrazowi medycznemu przypisuje się etykietę chory/zdrowy.

Dane uczące i model

Wyobraźcie sobie, że uczycie komputer rozpoznawania różnych obiektów, np. kotów i psów na zdjęciach. Potrzeba do tego dwóch rzeczy: samego komputera, który ma się uczyć (to nasz „uczeń"), i kogoś, kto mu pokaże, jak to robić (to nasz „nauczyciel").

W najczęściej używanej metodzie uczenia maszynowego i nadzorowanego wszystko wygląda trochę jak w szkole. „Nauczyciel" pokazuje „uczniowi" – komputerowi zdjęcia, na których są koty i psy, i mówi, które zwierzę jest które. Na przykład wskazuje na zdjęcie psa i mówi: „To jest pies", a potem pokazuje zdjęcie kota i mówi: „To jest kot". Następnie komputer stara się sam rozpoznać, czy na nowym zdjęciu jest kot czy pies. Jego celem jest, aby odpowiedź była jak najbardziej zbliżona do tego, co mu „nauczyciel" powiedział. Jeśli „nauczyciel" pokazał, że pewne zdjęcie przedstawia psa, to komputer też powinien to rozpoznać. W ten sposób, pokazując komputerowi wiele różnych zdjęć i mówiąc, co na nich jest, uczymy go odróżniać koty od psów. To właśnie nazywamy uczeniem maszynowym!

Uczenie maszynowe działa trochę jak nauka rozwiązywania równań matematycznych. Na początku komputer ma pewną funkcję, czyli zasadę, według której próbuje rozwiązać problem. Na przykład chcemy, aby nauczył się odróżniać zdjęcia kotów od zdjęć psów. W tym przypadku każde zdjęcie ma swoją etykietę – „kot" lub „pies", którą komputer musi odgadnąć.

Najistotniejszą częścią programu uczącego się jest fragment kodu (tzw. model), który zwraca „odpowiedzi" na zadane „pytania", w zależności od wewnętrznych ustawień (parametrów, tj. tablic liczb) owego programu. Dzięki temu, że te parametry zmieniają się ▸

Warto doczytać:
- A. Barr, E. A. Feigenbaum, *The Handbook of Artificial Intelligence*, Butterworth-Heinemann, 1981.
- S. J. Russell, P. Norvig, *Sztuczna inteligencja. Nowe spojrzenie*, tłum. A. Grażyński, Helion 2023.
- Y. LeCun, *A path towards autonomous machine intelligence*, version 0.9. 2, 2022-06-27, „Open Review" 2022, https://openreview.net/pdf?id=BZ5a1r-kVsf [dostęp: 6.12.2023].

w czasie uczenia, możliwy jest efekt uczenia, a więc model zaczyna coraz lepiej „odpowiadać" na zadawane mu „pytania" (np. coraz częściej prawidłowo określać, kiedy na zdjęciu jest kot, a kiedy pies). Warte podkreślenia jest, że jedyne, co się zmienia w modelu, to te wewnętrzne ustawienia – pewne liczby stają się trochę większe, inne trochę mniejsze. Tylko to! Nie zmienia się stopień skomplikowania modelu czy ciąg instrukcji procesora, który go realizuje (np. nigdy nie jest tak, że program czasami „pomyśli" dłużej, czasami krócej). Po zakończeniu nauki (po tzw. fazie treningu modelu), gdy parametry są już odpowiednio ustawione, również one nie zmieniają się już w ogóle i cały model można zapisać na dysku komputera.

Jak komputer „dowiaduje się" o błędach i czyni postępy?

Wyobraźcie sobie, że komputer, ucząc się odróżniać koty od psów na zdjęciach, czasami popełnia błędy. „Nauczyciel" sprawdza, czy odpowiedzi „ucznia" są poprawne, i mówi mu, kiedy ten ostatni się myli. Przypomina to sytuację, gdy nauczyciel w szkole sprawdza twoje zadania domowe.

W uczeniu maszynowym używa się czegoś, co nazywamy „funkcją straty". Możemy powiedzieć że jest to rodzaj miernika, który pokazuje, jak bardzo odpowiedzi komputera różnią się od prawidłowych. Jeśli wynik tej funkcji jest równy zero, oznacza to, że komputer nie popełnił żadnego błędu; im wynik jest wyższy, tym więcej błędów maszyna popełniła. Na podstawie tych informacji komputer uczy się, co poprawić. Przypomina to otrzymywanie wskazówek, co zrobić lepiej następnym razem. W procesie nauki komputer stopniowo udoskonala swoje „wewnętrzne ustawienia" (parametry), aby coraz lepiej odpowiadać na pytania. Podobnie jak w przypadku nauki gry na instrumencie – na początku robicie wiele błędów, ale z czasem gracie coraz lepiej. Tak samo komputer, ucząc się, z każdą pomyłką staje się coraz sprawniejszy w rozpoznawaniu, co jest na zdjęciu.

Oczywiście na początku nauki „uczeń" robi dużo błędów, ale powtarzając (być może wielokrotnie) przedstawiony proces, czyni postępy. Chociaż bywa i tak, że zredukuje liczbę błędów przez „wykucie na pamięć" bez „złapania" istoty problemu. By ograniczyć tego typu sytuacje, efekt nauki mierzy się z użyciem osobnego zestawu przykładów (zbioru testowego), niepokazywanych w czasie uczenia. To jak klasówka w szkole – zawiera podobne, lecz inne zadania niż te rozwiązywane na lekcji.

Sztuczne sieci neuronowe

Działanie sztucznych sieci neuronowych w uczeniu maszynowym bardzo przypomina zaawansowaną grę w „łączenie kropek", inspirowaną tym, jak funkcjonują ludzkie mózgi. Wyobraź sobie, że komputer ma sieć małych „neuronów" (są one jak minikomputery), które pracują razem, aby rozwiązywać zadania, np. rozpoznawać zdjęcia.

Każdy z tych „neuronów" robi coś bardzo prostego. Na przykład dostaje liczbę (reprezentującą fragment informacji, jak kolor piksela na zdjęciu) i przekształca ją w inną liczbę według prostych zasad. Te „neurony" są połączone w warstwy, a każda warstwa dostaje informacje od poprzedniej i przekazuje dalej do następnej. Dysponując wieloma warstwami „neuronów" ułożonych jedna na drugiej (stąd nazwa „głębokie uczenie"), komputer może nauczyć się rozpoznawać bardzo złożone wzorce. Na początkowych warstwach sieć uczy się identyfikować proste, lokalne cechy, takie jak krawędzie czy obszary o jednolitym kolorze. W miarę przechodzenia do kolejnych warstw sieć zaczyna rozpoznawać coraz bardziej złożone i abstrakcyjne elementy, jak np. struktura sierści czy części ciała zwierzęcia. W ten sposób kolejne warstwy sieci neuronowej stopniowo integrują te proste cechy w coraz bardziej kompleksowe wzorce, pozwalając na rozpoznawanie całościowych obrazów, takich jak sylwetka kota.

W praktyce oznacza to, że komputer, patrząc na wiele różnych zdjęć, uczy się, jakie cechy tworzą obraz „kota" czy „psa". Może to zrobić, bo każdy „neuron" i każda warstwa uczą się małych części obrazu, a potem łączą te informacje, by zrozumieć całość.

Wielkie modele językowe

Obecnie powszechnie dostępne są programy, które potrafią odpowiadać na nietrywialne pytania, spełniają dane im w języku naturalnym polecenia, tak jakby rozumiały. O tych tzw. wielkich modelach językowych (ang. *large language models* – LLMs) mówi się, że „nikt nie wie – nawet ich twórcy – jak one działają". Jak to jest możliwe i co to dokładnie znaczy?

Wydaje się, że jest to coś zupełnie innego niż choćby przedstawiony wcześniej przykład klasyfikacji obiektów na zdjęciach. Wyobraźcie sobie komputer, który ma nauczyć się samodzielnie pisać teksty, np. kontynuować zdanie. Jak się okazuje, to zadanie także można sprowadzić do rozpoznawania wzorców, podobnie jak uczenie się rozpoznawania obrazów.

W tym przypadku komputer uczy się przewidywać, jaka litera powinna pojawić się jako następna w danym tekście. Na przykład jeśli ma słowo filozofuj.e, to na podstawie tego, co już wie, próbuje zgadnąć, jaka litera powinna być dalej. Może to być „u", co daje słowo „filozofuj.eu". Komputer uczy się tego na podstawie ogromnej ilości tekstu z internetu. W każdym przykładzie ma fragment tekstu (jak „filozofuj.e") i literę, która występuje po nim (jak „u"). Przypomina to uczenie się słówek, gdy zna się ich kontekst. Analogicznie zachodzi „domyślanie się", jakie będzie następne słowo albo grupa słów.

W tym procesie komputer nie musi mieć dodatkowych etykiet wskazujących, co jest czym, ponieważ już ma tekst i kolejne litery, które występują naturalnie – w dostępnym mu otoczeniu informatycznym. Gdy uczenie komputera zakłada tego typu, automatycznie wygenerowane, dane, nazywamy je „samonadzorowanym", bo nie angażuje ono ludzi w proces etykietowania kolejnych przykładów uczących.

Implementacja – w informatyce: przełożenie jakiegoś algorytmu (wyrażonego np. słownie lub narysowanego na schemacie blokowym) na program komputerowy (napisany w jakimkolwiek języku programowania), który komputer może wykonać.

Podczas tego uczenia komputer tworzy kontynuację tekstu, bazując na tym, co już wie. Każda nowa litera, którą dodaje, staje się częścią tekstu, który już napisał, i tak dalej, aż do końca zdania lub większego fragmentu tekstu.

Wytrenowany duży model językowy robi tylko to: zwraca (przedstawia nam) prawdopodobną kontynuację tekstu na podstawie statystycznych zależności odkrytych w zbiorze uczącym (dostępnych danych). Czy jest to tekst prawdziwy, czy nie, piękny czy nie, użyteczny czy nie, logicznie spójny czy nie – wszystko to jest pochodną tego, co znajdowało się w zbiorze uczącym szczegółów procesu uczenia i każdej literki początkowego kontekstu. Z tego powodu, by odpowiedzi były jak najczęściej prawdziwe, niekrzywdzące i najbardziej pomocne, modele językowe są dotrenowywane na znacznie mniejszych zbiorach specjalnie przygotowanych przykładów par: pytanie – „jakościowa" odpowiedź.

Nietransparentność sieci neuronowych

Sieci neuronowe i ich działania są niezwykle złożone. Generowanie tekstu oparte jest na potężnej liczbie parametrów, jakie „ustawiły się" w czasie uczenia. Nie sposób zinterpretować tych liczb, nie wiadomo, które z nich co konkretnie oznaczają, na jaki aspekt znaczenia słów poszczególne neurony reagują dużymi aktywacjami. Twórcy wielkich modeli językowych znają każdy najmniejszy szczegół ich implementacji i treningu, ale również są bezradni w „rozczytywaniu" dokładnego przebiegu i sensu wszystkich obliczeń, które określają ostateczną „wypowiedź" modelu. Powyższy opis pracy systemów SI jest niezwykle uproszczony, mam jednak nadzieję, że umożliwia pewne zrozumienie rządzących nimi „mechanizmów". ■

Pytania do tekstu

1. Czym różni się wąska SI od ogólnej sztucznej inteligencji?
2. Na czym polega tzw. głębokie uczenie?
3. Czym są wielkie modele językowe?

Czy sztuczna świadomość jest możliwa?

Robert Poczobut

Profesor UwB, kierownik Zakładu Epistemologii i Kognitywistyki w Instytucie Filozofii UwB. Autor czterech książek, przeszło siedemdziesięciu artykułów naukowych, redaktor licznych prac zbiorowych. Interesuje się pograniczem filozofii umysłu i kognitywistyki, w szczególności naturą jaźni, relacją między świadomymi i nieświadomymi procesami poznawczymi, neurofilozofią oraz hybrydowymi systemami poznawczymi. Hobby: podróże, muzyka, malarstwo, gry logiczne.

W filozofii sztucznej inteligencji stawiamy pytanie o możliwość istnienia umysłów niebiologicznych, zbudowanych na bazie innego substratu (budulca) niż struktury węglowo-białkowe. Chcielibyśmy wiedzieć, czy świadomość u swych podstaw musi być organiczna, czy też jej istnienie jest w jakimś sensie niezależne od substratu.

Słowa kluczowe: świadomość fenomenalna, sztuczna świadomość, zasada niezależności od substratu, organizacja funkcjonalna, funkcjonalizm

Możliwości i wartości

Dotychczas nikt nie wykazał, że istnienie sztucznej świadomości jest niemożliwe. Nie przedstawiono dowodu, że jest to pojęcie wewnętrznie sprzeczne; nie wykazano, że świadomość maszynową wykluczają prawa przyrody. Dopóki takie dowody nie zostaną podane, możemy racjonalnie wierzyć w możliwość istnienia świadomej sztucznej inteligencji (SI). Tym bardziej że istnieją przesłanki, które taką możliwość uwiarygodniają. Postęp w dziedzinie SI stopniowo, ale względnie szybko w porównaniu z ewolucją biologiczną obejmuje coraz bardziej wyrafinowane funkcje poznawcze. Zgodnie z metaforą Hansa Moraveca krajobraz ludzkich kompetencji jest stopniowo zatapiany przez szybko podnoszącą się wodę, która symbolizuje ekspansję sztucznej inteligencji. Kolejne zdolności poznawcze, jeszcze niedawno zarezerwowane wyłącznie dla ludzi, stają się udziałem maszyn. Możliwe, że w niedalekiej przyszłości zostaną „zatopione" również kompetencje związane ze świadomością.

Gdyby rzeczywiście sprawy szły w tym kierunku, a wydaje się, że tak właśnie jest, moglibyśmy pytać, czy warto dążyć do wytworzenia świadomych maszyn. Czy do realizacji ludzkich potrzeb nie wystarczą inteligentne narzędzia pracujące w trybie zombie (bez świadomości)? Powstanie świadomych maszyn wiązałoby się z nadaniem im statusu moralnego i prawnego. Nie traktowalibyśmy ich jak narzędzi, lecz uznalibyśmy za osoby. Niektórych taka perspektywa przeraża, innych mobilizuje do eksploracji wciąż niezrealizowanych możliwości. Niewykluczone, że wraz z powstaniem sztucznej świadomości maszyny rozwiną także jakąś formę emocji, w swoich decyzjach będą zaś kierować się zaszczepioną lub rozwiniętą hierarchią wartości. ▶

Hans Peter Moravec (ur. 1948) – futurolog i transhumanista, badacz sztucznej inteligencji; profesor w Instytucie Robotyki na Uniwersytecie Carnegie Mellon w Pittsburghu; w swoich pracach analizuje konsekwencje ewolucji inteligencji robotów.

Wówczas różnica między inteligencją naturalną a sztuczną przestanie być wyraźna. Jak twierdzi Frank Wilczek, laureat Nagrody Nobla w dziedzinie fizyki:

> wszelka inteligencja jest maszynowa, a inteligencję naturalną od sztucznej odróżnia nie to, czym są, ale jak są zrobione (2020, s. 85).

Dwa oblicza świadomości – funkcjonalne i fenomenalne

Niektórzy, mówiąc o świadomości, mają na uwadze zdolność do automonitoringu, samodzielnego kontrolowania swojego stanu. Dysponują nią organizmy biologiczne i maszyny o hierarchicznych architekturach z mechanizmem informacyjnych sprzężeń zwrotnych[9]. Działanie takich maszyn wiąże się z dostępem do informacji o ich stanach wewnętrznych. Nie ulega wątpliwości, że niektóre z istniejących maszyn dysponują zdolnością do automonitoringu i kontroli zachowania. Filozofowie nazywają takie zdolności „świadomością funkcjonalną" lub „świadomością dostępu". Jednak słusznie zwraca się uwagę, że nie uwzględniają one istotnych cech świadomości biologicznej jak doznawanie wrażeń i przeżywanie stanów wewnętrznych. Maszyny z wbudowanym systemem automonitoringu i kontroli zachowania, lecz pozbawione przeżyć świadomych, są systemami typu zombie.

W dyskusjach na temat świadomości wyróżnia się też pojęcie świadomości fenomenalnej. Polega ona na zdolności do posiadania subiektywnych przeżyć – mogą to być proste doznania zmysłowe, jak odczucie bólu czy doznanie zieleni, lub złożone przeżycia świadome towarzyszące emocjom, myśleniu czy aktywności twórczej. Powstaje pytanie, czy maszyny mogą mieć przeżycia świadome, choćby w postaci prostych doznań zmysłowych. Opinie filozofów i naukowców są w tej kwestii podzielone. Aby konkluzywnie odpowiedzieć na to pytanie, musielibyśmy wiedzieć, na mocy jakich mechanizmów w organizmach biologicznych powstają przeżycia świadome i czy takie mechanizmy można odtworzyć w systemach sztucznych.

Sprzężenie zwrotne – oddziaływanie skutku działania na jego przyczynę.

Ilustracja: ChatGPT

Substrat świadomości i organizacja funkcjonalna

Wszystkie systemy, którym przypisujemy przeżycia świadome, są organizmami biologicznymi. Z tego powodu świadomość fenomenalna wydaje się nam szczególnie mocno związana z substratem biochemicznym. Inne procesy poznawcze, jak uczenie się czy pamięć, skłonni jesteśmy traktować jako niezależne od substratu (można je zrealizować w różnych substratach fizycznych). Zdaniem Maxa Tegmarka[9] do realizacji procesu poznawczego konieczny jest nie określony rodzaj materii, lecz jej organizacja funkcjonalna. Pamięć, obliczanie, uczenie się i inteligencja są niezależne od fizycznych szczegółów substratu, dzięki czemu możliwe jest istnienie systemów SI i maszyn realizujących procesy poznawcze. Jak pisze Tegmark:

> Sprzęt jest materią, a oprogramowanie – wzorcem. Niezależność obliczeń od substratu oznacza, że możliwe jest istnienie sztucznej inteligencji: inteligencja nie wymaga ciała, krwi ani atomów węgla (2019, s. 94).

Czy podobnie jest ze świadomością fenomenalną?

David Chalmers[9], zastanawiając się, w jaki sposób systemy fizyczne wywołują świadomość fenomenalną, również zwraca uwagę, że kluczowa jest ich organizacja funkcjonalna, a nie substrat, z którego są zrobione. Dotyczy to także przeżyć świadomych wytwarzanych przez nasze mózgi. To, że mózgi wytwarzają świadomość, nie jest dziwniejsze od tego, że mogłyby

Max Tegmark (ur. 1967) – profesor fizyki, kosmolog. Pracuje w Massachusetts Institute of Technology. Jego badania dotyczą wykorzystaniu sztucznej inteligencji w fizyce i fizyki w badaniach nad sztuczną inteligencją.

David Chalmers – australijski filozof i kognitywista, specjalizujący się w filozofii umysłu i filozofii języka. Jest profesorem filozofii Uniwersytetu Nowojorskiego i jednym z dyrektorów w Center for Mind, Brain, and Consciousness. Autor przełożonej na język polski bestsellerowej książki *Świadomy umysł* (2010).

Sztuczna inteligencja jako podróż w Nieznane

Piotr Kulicki

Profesor filozofii, dyrektor Instytutu Filozofii KUL i Center for Human Oriented Artificial Intelligence na KUL-u. Od strony naukowej zajmuje się logiką i ontologią formalną oraz ich zastosowaniami na potrzeby sztucznej inteligencji. Z czasem coraz więcej uwagi przykłada do czerpania satysfakcji z doświadczeń życia codziennego.

O sztucznej inteligencji i jej spektakularnych postępach napisano w ostatnich latach bardzo wiele, także w „Filozofuj!", i właściwie trudno coś nowego tu dodać. Dlatego chciałbym zwrócić uwagę raczej na to, jakie są reakcje ludzi na jej postępujący rozwój. Myślę, że jest to część szerszego zjawiska – reakcji ludzi na to, co nieznane, niezrozumiałe czy nowe. Reakcje te odnotować można w odbiorze świata od wieków i w zasadzie od wieków się nie zmieniają.

Słowa kluczowe: sztuczna inteligencja, nieznane, superinteligencja

Czym może być Nieznane?

Sięgnijmy do tak lubianych przez filozofów, i nie tylko przez nich, początków starożytnej Grecji. Wtedy znany świat ograniczał się do zamieszkanych przez Swoich, rozsianych po wybrzeżach Morza Śródziemnego greckich miast. To, co na zewnątrz, pełne było groźnych potworów, czasem podstępnych jak syreny, wabiące urzekającym śpiewem podróżnych, by następnie ich zabijać, a bogowie wkraczali w losy ludzi kierowani sobie tylko znanymi motywami. Były też poza znanym światem rzeczy niezmiernie atrakcyjne. W oddali śmiałkowie znaleźć mogli bogactwo, takie jak złote runo, które było w stanie szybko odmienić ich los, a gdzieś za Słupami Herkulesa leżała opisana przez Platona Atlantyda, w której ludzie potrafili żyć szczęśliwie, zachowując idealny porządek społeczny. Tam więc, gdzie zaczynało się Nieznane, obawy przed najgorszym spotykały się z nadzieją na lepsze życie.

Z lat mojego dzieciństwa pamiętam inne Nieznane. Przybrało ono postać Niezidentyfikowanych Obiektów Latających (UFO) i innych przejawów kontaktów z obcymi cywilizacjami bądź istotami pozaziemskimi. Pozostają z tego czasu odnoszące się do tematu dzieła kultury, szczególnie tej popularnej. Dużo wcześniej inwazja Marsjan pojawiła się w stylizowanym na reportaż słuchowisku radiowym *Wojna światów*, nadanym w Nowym Jorku w Halloween roku 1938. Ludzkość przetrwała, ale tylko dlatego, że Marsjanie okazali się nieodporni na ziemskie mikroby. Przełom lat 70. i 80. XX w. obfituje w utwory o zbliżonej tematyce. W filmie *Obcy. Ósmy pasażer Nostromo* tytułowy obcy jest podobny do starożytnej hydry. Nie próbuje kontaktować się z ludźmi, nie dowiadujemy się o jego inteligencji czy podobieństwie do nas – jest za to na pewno śmiertelnym zagrożeniem. *Bliskie spotkania trzeciego stopnia* przedstawiają wizję mniej złowieszczą. Obca cywilizacja nie jest jawnie wroga. Obawa łączy się z fascynacją i nadzieją na poprawę świata, którą może przynieść kontakt z bardziej rozwiniętymi przybyszami z kosmosu. Podobnie jest w nieco późniejszym filmie *Kontakt* (1997), gdzie spotkanie z obcą cywilizacją jest dla głównej bohaterki przeżyciem niemalże mistycznym, pozwalającym na nowo odkryć sens życia, choć forma tego spotkania jest dla widza zaskakująca.

Niepoznana sztuczna inteligencja

Dla większości ludzi, którzy zajmują się aktualnie tworzeniem systemów określanych mianem sztucznej inteligencji i świadome są mechanizmów uczenia maszynowego, jest naturalne, że są one tylko narzędziami służącymi do realizacji postawionych im konkretnych zadań. Zapowiadane przez niektórych wizjonerów takich jak Ray Kurzweil pojawienie się superinteligentnych maszyn wydaje się nierealne lub przynajmniej odległe. Jednakże rzeczywisty postęp w realizacji zadań, które dotąd dostępne były tylko dla ludzi, jest imponujący. Sprawia to, że możemy zwątpić w to, co wydaje się oczywiste. Głośnym echem odbiła się informacja, że inżynier oprogramowania z firmy Google, Blake Lemoine, stwierdził, iż program, nad którym pracuje, jest świadomą, mogącą mieć duszę istotą i powinny przysługiwać jej odpowiednie do tego statusu prawa.

Staje się coraz bardziej jasne, że nie wiemy, jak daleko w stronę uznawanych pierwotnie za wyłącznie ludzkie umiejętności może zajść technologia. Posłużmy się tu obrazem, który w roku 1997 przedstawił Hans Moravec. Bogactwo ludzkich kompetencji ujął w metaforę krajobrazu, gdzie to, co najbliższe nam, najbardziej ludzkie i jednocześnie najtrudniejsze do naśladowania przez automaty (interakcje społeczne, wyrażanie się poprzez sztukę, ale też koordynacja wzrokowo-ruchowa, naturalne poruszanie się) znajduje się na wzniesieniach górskich, a kłopotliwe dla nas, ale łatwe w automatyzacji czynności (pamięciowe opanowanie materiału szkolnego, wykonywanie obliczeń arytmetycznych) znajdują się na nizinach, natomiast szereg pośrednich umiejętności (dowodzenie twierdzeń, gra w szachy) znajduje się na wzgórzach pomiędzy nimi. Rozwój narzędzi sztucznej inteligencji jest jak powódź, w której woda przejmuje od nas kolejne kompetencje, poczynając od tych, które położone są najniżej. Dwadzieścia lat później Max Tegmark w książce *Życie 3.0* przedstawił stan mu aktualny na rysunku podobnym do zamieszczonego poniżej. ▶

Warto doczytać:

- P. Bołtuć, *BICA jako szansa stworzenia świadomych maszyn*, „Przegląd Filozoficzny. Nowa Seria" 2013, nr 2, s. 185–196.
- D. Chalmers, *Świadomy umysł*, tłum. M. Miłkowski, Warszawa 2010.
- S. Dehaene, *Świadomość i mózg. Odczytywanie kodu naszych myśli*, tłum. D. Rossowski, Kraków 2023.
- P. Haikonen, Robot Brains: *Circuits and Systems for Conscious Machines*, Chichester 2007.
- R. Kurzweil, *How to Create a Mind*, New York 2012.
- S. Schneider, *Świadome maszyny. Sztuczna inteligencja i projektowanie umysłu*, tłum. J. Bednarek, Warszawa 2021.
- M. Tegmark, *Życie 3.0. Człowiek w erze sztucznej inteligencji*, tłum. T. Krzyształ, Warszawa 2019.
- F. Wilczek, *Jedność inteligencji*, [w:] *Człowiek na rozdrożu. Sztuczna inteligencja: 25 punktów widzenia*, J. Brockman (red.), Gliwice 2020, s. 85–95.

ją wytwarzać układy sztuczne. Jak pisze Chalmers:

> „Dowolny system o organizacji funkcjonalnej odpowiedniego rodzaju jest świadomy bez względu na to, z czego się składa (2010, s. 513).

Zdaniem funkcjonalistów fizyczna implementacja (zob. na s. 8 tego numeru) odpowiedniej architektury obliczeniowej wystarcza do osiągnięcia organizacji funkcjonalnej umożliwiającej wytworzenie świadomości fenomenalnej. Systemy świadome są układami fizycznymi o strukturze przyczynowej odzwierciedlającej formalną strukturę obliczeń. Sama formalna struktura obliczeń, czyli oprogramowanie jako obiekt abstrakcyjny, nie wystarcza do powstania świadomości. Dopiero jej fizyczna implementacja w postaci układu konkretnych procesów fizycznych powiązanych przyczynowo umożliwia powstanie świadomości fenomenalnej. Innymi słowy, to organizacja funkcjonalna układu określa podstawowy mechanizm świadomości fenomenalnej. Zadaniem neurobliczeniowej teorii świadomości jest odkrycie organizacji funkcjonalnej mózgu, która odpowiada za wytwarzanie świadomości. Zadaniem mocnej wersji sztucznej inteligencji (zakładającej jej świadomość fenomenalną) jest odtworzenie tej organizacji w systemach sztucznych.

Pytania bez odpowiedzi

W dalszym ciągu nie wiemy w szczegółach, jaka organizacja funkcjonalna umożliwia powstanie świadomości fenomenalnej w mózgu. Nie wiemy również, czy organizacja funkcjonalna mózgu jest możliwa do odtworzenia w środowisku sztucznym. W odróżnieniu od funkcjonalistów zwolennicy podejścia substratowego utrzymują, że tylko układy biologiczne, zbudowane na bazie chemii organicznej, ostatecznie zaś na bazie związków węgla, odznaczają się taką plastycznością i dynamiką, które umożliwiają wytworzenie świadomości. Organizacja funkcjonalna układu świadomego nie jest niezależna od jego substratu. Związki krzemu lub innych pierwiastków nie mają odpowiednich cech strukturalnych umożliwiających wytworzenie świadomych umysłów.

Funkcjonaliści uznają świadomość za cechę emergentną[1] odpowiednio złożonych układów fizycznych – cechę niezależną od szczegółowych własności jej substratu. Wiążąc świadomość z organizacją funkcjonalną, pozostawiają otwartą możliwość istnienia sztucznej świadomości. Współcześnie stanowisko to ma wpływowych przedstawicieli wśród filozofów i naukowców (m.in. wspomniani D. Chalmers i M. Tegmark, a także Giulio Tononi i Stanislas Dehaene). Jak pisze Tegmark:

> „Świadomość jest zjawiskiem fizycznym odczuwanym niefizycznie, ponieważ ma taki sam charakter jak fale i obliczenia; ma własności niezależne od swojego konkretnego podłoża fizycznego. Wynika to logicznie z koncepcji świadomości jako informacji […]: jeśli świadomość to sposób, w jaki odczuwa się informację, gdy jest ona przetwarzana w określony sposób, to musi być niezależna od podłoża; liczy się struktura przetwarzania informacji, a nie struktura materii dokonującej tego przetwarzania. Innymi słowy, świadomość jest podwójnie niezależna od podłoża (2019, s. 388–389).

Pytania do tekstu

1. Czy możemy racjonalnie wierzyć w możliwość istnienia świadomej sztucznej inteligencji?
2. Czym jest świadomość funkcjonalna, a czym fenomenalna?
3. Czym różnią się funkcjonaliści od zwolenników podejścia substratowego w kwestii możliwości zaistnienia świadomości?

Emergencja (od łac. emergo, emergere – „wyłaniam się", „powstaję") – pojawienie się jakiejś nowości bytowej na bazie występujących już elementów wchodzących w skład pewnej całości.

Grafika na podstawie ilustracji: M. Tegmark, *Life 3.0: Being Human in the Age of Artificial Intelligence*, New York 2017, s. 53.

Od tego czasu minęło sześć lat, a zauważyć można, że sztuczna inteligencja wchodzi już w miejsca pozostające na obrazie Tegmarka daleko od wody: tworzy z powodzeniem obrazy, które można uznać za dzieła sztuki, komponuje utwory muzyczne, pisze książki, dowodzi twierdzeń, używając technik matematycznych wcześniej ludziom nieznanych, pisze programy komputerowe, sprawnie tłumaczy, streszcza i rozbudowuje teksty. Coraz trudniej znaleźć te miejsca, w których jako ludzie pozostajemy wyjątkowi. Sprawia to, że przyszłość sztucznej inteligencji rzeczywiście trzeba uznać za wielką tajemnicę – znane nam od wieków Nieznane.

Skrajne reakcje na Nieznane są najbardziej typowe

Do sławnych osób, które wyraźnie artykułowały obawy związane z rozwojem inteligentnych technologii, należał fizyk Stephen Hawking. W 2016 r. twierdził, że problemem dla nas będzie wyjątkowa sprawność przyszłych superinteligentnych systemów, gdy cele, które będą one realizować, nie będą zbieżne z naszymi. Podał sugestywny przykład, zwracając się do swojego rozmówcy w następujący sposób: „Prawdopodobnie nie jesteś osobą, która nienawidzi mrówek i depcze je ze złości, ale jeśli jesteś odpowiedzialny za projekt hydroelektrowni, a w regionie znajduje się mrowisko, które musi zostać zalane, to tylko źle dla mrówek. Nie pozwólmy, aby ludzkość znalazła się w sytuacji tych mrówek" (netoteka). Z kolei Geoffrey Hinton, zwany przez niektórych ojcem chrzestnym sztucznej inteligencji, odszedł w maju tego roku z Google, wyjaśniając, że zrezygnował z pracy, gdyż chce bez skrępowania mówić o zagrożeniach związanych ze sztuczną inteligencją. Dodał, że po części żałuje wkładu, który wniósł w rozwój tego typu technologii. Niepokoją go zwłaszcza zalew dezinformacji i głębokie zmiany na rynku pracy, szczególnie w sytuacji, w której w rozwoju sztucznej inteligencji przodują globalne korporacje. Wskazuje też na bliżej nieokreślone *zagrożenie egzystencjalne*, które pojawi się wraz z powstaniem prawdziwej inteligencji cyfrowej.

Niemniej wielu znawców tematu, w tym ważni współtwórcy obecnych sukcesów technologii tacy jak Jürgen Schmidhuber czy Yann LeCun, przyjmuje zdecydowanie odmienną postawę wobec nieznanej w szczegółach przyszłości inteligentnych technologii. Znamienny dla tego podejścia jest tytuł artykułu z „Rzeczpospolitej" autorstwa Zbigniewa Gajewskiego: *Sztuczna inteligencja rozwiąże problemy globu*. Wśród wspomnianych problemów znajduje się wszystko to, co danego autora niepokoi we współczesnym świecie. Wskazywane są: zbyt powolny rozwój nauki, kryzys klimatyczny, pandemie, głód panujący w wielu miejscach na świecie i jednocześnie marnotrawstwo dóbr. Wszystkie one mają być rozwiązane przez właściwe wykorzystanie sztucznej inteligencji.

Czy można się przygotować na Nieznane?

Osobiście radzę zachować umiar. Jestem bardzo ciekawy tego, co przyniesie przyszłość sztucznej inteligencji, ale zarówno najpoważniejsze obawy, jak i wielkie nadzieje uważam za przesadne. Na pewno wiele się zmieni i warto być na to gotowym. Jestem też przekonany, że nikt teraz nie jest w stanie przewidzieć, co konkretnie nas czeka. Możemy więc zrobić tylko to, co jest niezależne od sposobu, w jaki Nieznane nas zaskoczy: postarać się lepiej zrozumieć świat i ludzi, racjonalnie odczytywać sens komunikatów, które otrzymujemy, oraz intencje, które za nimi się kryją, a także rozumieć samego siebie, własne potrzeby i oczekiwania. W tym kontekście warto przytoczyć jeden z postulatów dotyczący edukacji przyszłości, który przedstawił Jacques Attali, francuski ekonomista, doradca prezydenta Francji François Mitterranda: od czwartego roku życia nauczać dzieci filozofii. ■

Netoteka:

■ https://www.reddit.com/r/science/comments/3nyn5i/science_ama_series_stephen_hawking_ama_answers/?rdt=56321

Pytania do tekstu

1. Jak rozumiesz metaforę porównującą rozwój narzędzi sztucznej inteligencji do powodzi?
2. Jakimi obawami na temat sztucznej inteligencji dzielił się z nami Stephen Hawking w 2016 r.?
3. Jakie problemy ma zdaniem Zbigniewa Gajewskiego rozwiązać sztuczna inteligencja?

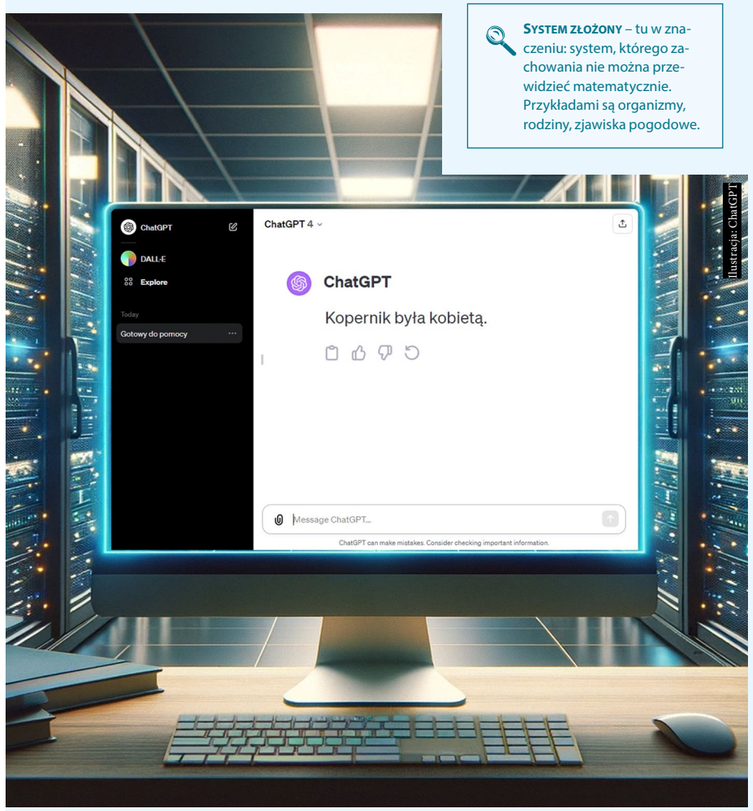
Dlaczego ChatGPT nigdy nie będzie rządzić światem

Barry Smith

Jeden z najczęściej cytowanych współczesnych filozofów; profesor filozofii, informatyki biomedycznej i informatyki na Uniwersytecie w Buffalo, dyrektor the National Center for Ontological Research (Narodowego Centrum Badań Ontologicznych). Jego strategia wykorzystania ontologii do rozwiązywania rzeczywistych problemów w zarządzaniu danymi jest pierwszą koncepcją filozoficzną, która została zatwierdzona jako standard ISO (ISO/IEC:21838-2).

Jobst Landgrebe i ja w naszej najnowszej książce *Why Machines Will Never Rule the World* (*Czemu maszyny nigdy nie będą rządzić światem*) argumentujemy, że wysiłki wielu osób ze społeczności sztucznej inteligencji zmierzające do stworzenia ogólnej sztucznej inteligencji (*artificial general intelligence*, AGI) są skazane na niepowodzenie. Mówiąc w tym przypadku o AGI, mamy na myśli maszynę, która wykazywałaby zdolności poznawcze równoważne lub nawet przewyższające ludzkie.

Słowa kluczowe: AGI, ChatGPT



Ilustracja: ChatGPT

SYSTEM ZŁOŻONY – tu w znaczeniu: system, którego zachowania nie można przewidzieć matematycznie. Przykładami są organizmy, rodziny, zjawiska pogodowe.

Nasz argument za niemożliwością istnienia ogólnej sztucznej inteligencji ma następującą postać:

1. Analizujemy właściwości systemów złożonych (złożonych w innym sensie niż systemy logiczne), takich jak ziemski system pogodowy lub system ruchu drogowego w Stambule.

2. Wykazujemy, że istnieją poważne ograniczenia naszej zdolności do matematycznego przewidywania zachowań tego rodzaju systemów.

3. Pokazujemy, że ograniczenia te determinują również zdolność komputerów do dokonywania takich przewidywań.

Nasz wniosek co do niemożliwości istnienia ogólnej sztucznej inteligencji wynika z faktu, że wszystkie systemy organiczne – w tym ludzki system neurologiczny – są systemami złożonymi (w powyższym rozumieniu). AGI wymagałaby komputerów zdolnych do przewidywania zachowań istot ludzkich z dużą niezawodnością. Zdolność ta jest niezbędna, jeśli na przykład maszyna ma inteligentnie uczestniczyć w rozmowie z ludźmi. I to właśnie brak tej zdolności wyjaśnia problemy, jakie napotykamy podczas rozmów telefonicznych z komputerami naszych banków. Maszyny te wciąż wykazują mizerny poziom wydajności, nawet ▶

SYSTEM LOGICZNY – tu w znaczeniu: skomplikowany system, który zachowuje jednak cechę prostoty – w tym sensie, że jego zachowanie można przewidzieć za pomocą maszyny. To, że jest to możliwe, jest w rzeczywistości sprawą trywialną, ponieważ sama czynność *dostarczania wyników* przez ChatGPT to z matematycznego punktu widzenia przypadek *przewidywania tych wyników* (zob. też Thurner, Hanel, Klimek 2018).

Probabilistyczny – od łac. *probabilis* – „prawdopodobny"; odwołujący się do teorii prawdopodobieństwa.

Wektor binarny – zbiór liczb wyrażonych jako cyfry binarne uporządkowane liniowo. Na przykład wektor binarny <10100, 1000110> o długości dwa może przechowywać pomiary cali deszczu zebrane w danym dniu przez deszczomierz i średnią temperaturę tego dnia w stopniach Fahrenheita (w tym przypadku odpowiednio 20 cali i 70°).

po 50 latach prób nauczenia ich przez inżynierów AI symulowania podobnych rozmów. Takie wyrafinowane kompetencje w komunikowaniu się z ludźmi – np. z kontrolerami armii, dostawcami amunicji i tak dalej – byłyby oczywiście niezbędne, gdyby maszyny miały przejąć kontrolę nad światem.

Nasze argumenty sugerują, że komputery zawsze będą ograniczone do korzystania z algorytmów będących w stanie przewidzieć zachowania jedynie prostych systemów, takich jak laptopy czy linie montażowe w fabrykach. Z tego powodu zawsze będą charakteryzować się tym, co nazywamy „wąską sztuczną inteligencją", a tym samym zawsze będzie im brakować ogólnej inteligencji istot ludzkich.

Wejdź do programu ChatGPT

Przy powyższym wniosku, że ogólna sztuczna inteligencja nie może istnieć, obstawaliśmy przynajmniej wówczas, kiedy nasza książka została opublikowana we wrześniu 2022 r. Ale potem, w listopadzie 2022 r., nowy i rewolucyjny rodzaj sztucznej inteligencji został wypuszczony na świat w postaci programu ChatGPT. ChatGPT był inny. Stanowił pierwszy przykład systemu sztucznej inteligencji, który był łatwo dostępny dla zwykłych ludzi i zapewniał godziny niezakłóconej stymulacji. Ponadto zapewniał szereg różnych rodzajów usług, które mogą być wartościowe np. dla organizacji komercyjnych. ChatGPT był również, jak się wydaje, przykładem ogólnej sztucznej inteligencji, ponieważ mógł w swój sposób reagować na dowolne zapytania (Agüera y Arcas, Norvig 2023).

Bardzo szybko jednak rosnąca społeczność użytkowników programu ChatGPT stała się świadoma pewnych nieoczekiwanych zachowań, zwanych „halucynacjami", ze strony tego algorytmu. Jest on najwyraźniej statystycznie przygotowany do unikania zwrotów takich jak „nie wiem" w swoich wynikach. Zamiast tego często wymyśla własne odpowiedzi na pytania, na które nie zna odpowiedzi. Krótko mówiąc, wykazuje zachowanie, które, gdyby było prezentowane przez człowieka, byłoby określone jako „kłamanie" (zob. netoteka).

Jak działa ChatGPT

Aby zrozumieć, co się dzieje z tym „kłamaniem" AI, musimy najpierw uznać, że ChatGPT w rzeczywistości nic nie wie. Kiedy udziela odpowiedzi *O* na pytanie *P*, to nie dlatego, że wie, że jakaś teza *O* jest prawdziwa.

ChatGPT to zestaw algorytmów w formie oprogramowania, zbudowany na podstawie probabilistycznego modelu GPT, który służy do przetwarzania języka naturalnego. Kiedy zatem wprowadzamy jakiś monit *P*, algorytm ChatGPT przewiduje, że biorąc pod uwagę dane, które zostały użyte jako zestaw treningowy, dana sekwencja tokenów (w przybliżeniu: sylab) *O* jest (w przybliżeniu) następną najbardziej prawdopodobną sekwencją tokenów, biorąc pod uwagę *P* jako punkt wyjścia. (Należy zauważyć, że „przewidywanie" oznacza tutaj, że algorytm przedstawia *O* jako dane wyjściowe, biorąc pod uwagę *P* jako dane wejściowe). Po drugie, musimy zdać sobie sprawę, że ChatGPT działa zawsze tylko przez czerpanie z określonego (bardzo dużego) zbioru danych, na którym został przeszkolony. Oznacza to, że jest również przykładem wąskiej sztucznej inteligencji. Dzieje się tak, ponieważ jego algorytm nie odnosi się do rzeczywistego świata, w którym żyjemy, ani do wielu nakładających się na siebie złożonych systemów, z których nasz świat się składa. Odnosi się raczej do pewnego abstrakcyjnego symulakrum (kopii nieposiadającej oryginału) świata, symulakrum, które jest dokładnie określone przez duży, ale skończony zestaw danych, na których algorytm został wyszkolony, w podobny sposób, w jaki świat gry wideo jest określony przez jej oprogramowanie. W przypadku programu ChatGPT te dane treningowe zostały zdefiniowane (z grubsza) przez zawartość internetu w danym dniu w przeszłości (stąd użycie modyfikatora „Od ostatniej aktualizacji mojej wiedzy we wrześniu 2021 r." – aby uzasadnić, że nie udziela odpowiedzi na pytania dotyczące nowszych wydarzeń).

Algorytm, zdefiniowany w procesie szkolenia na podstawie danych dostępnych w 2021 r., jest funkcją matematyczną, która przyjmuje jako dane wejściowe wektory binarne kodujące polecenia *P* i wyprowadza wektory binarne kodujące odpowiedzi *O*. Możemy myśleć o tej funkcji jako o bardzo, bardzo długim równaniu wielomianowym (z ok. 1,5 miliarda parametrów). Stąd też szeroki zakres tematów, na które może udzielać odpowiedzi. Jednocześnie jednak zdolności matematyczne wymagane przez to równanie są nadal bardzo proste – ponieważ równanie, podobnie jak wszystkie obliczalne algorytmy, musi być możliwe do rozwiązania przy użyciu tylko bardzo prostej matematyki maszyny Turinga. (ChatGPT jest pod tym względem porównywalny do Tłumacza Google. Ten ostatni może być stosowany do wielu języków używanych do bardzo wielu tematów. Google Translate jest jednak przykładem wąskiej sztucznej inteligencji. Podobnie jak ChatGPT działa tylko dla bardzo prostych i niezmiennych światów, które są zdefiniowane przez dane, na których dana wersja oprogramowania została przeszkolona).

Na koniec zauważmy, że sytuacja nie poprawi się na korzyść AGI nawet wtedy, gdy uda się rozwiązać problem „halucynacji" programu ChatGPT. Nawet jeśli przyszłe systemy sztucznej inteligencji będą w stanie zmniejszyć stopień, w jakim generują „halucynacje", to nadal świat, w odniesieniu do którego każdy taki system udziela odpowiedzi, będzie niezmiennym światem zdefiniowanym przez system logiczny. Nie przypomina więc w niczym świata, w którym przez miliony lat ewoluowała ludzka inteligencja. ■

Tłumaczenie: Piotr Biłgorajski

Pytania do tekstu

1. Co jest istotą argumentu Barry'ego Smitha i Jobsta Landgrebe'a przeciw możliwości istnienia ogólnej sztucznej inteligencji (AGI)?
2. Czym są „halucynacje" programu ChatGPT?
3. Czy możemy powiedzieć, że AI coś wie?
4. Czy ChatGPT jest również przykładem wąskiej sztucznej inteligencji?

Warto doczytać:

- J. Landgrebem, B. Smith, *Why Machines Will Never Rule the World. Artificial Intelligence Without Fear*, Abingdon 2022.
- S. Thurner, R. Hanel, P. Klimek, *Introduction to the theory of complex systems*, Oxford 2018.
- B. Agüera y Arcas, P. Norvig, *Artificial General Intelligence Is Already Here*, „Noema Magazine", October 10, 2023, https://www.noemamag.com/artificial-general-intelligence-is-already-here/ [dostęp: 30.11.2023].

Netoteka:

- https://www.youtube.com/playlist?list=PLyngZgII3WThR6bptbvVACoCxn_mUCgCq

Czym zajmuje się etyka sztucznej inteligencji?

Artur Szutta

Filozof, pracownik Uniwersytetu Gdańskiego, specjalizuje się w filozofii społecznej, etyce i metaetyce. Jego pasje to przyrządzanie smacznych potraw, nauka języków obcych (obecnie węgierskiego i chińskiego), chodzenie po górach i gra w piłkę nożną.

Etyka sztucznej inteligencji (AI) to dziedzina etyki szczegółowej. Jej zakres i treść wyznaczają pytania, które stawiają etycy, badając zagadnienie sztucznej inteligencji. Oto główne kwestie oraz idee, jakimi zajmują się etycy AI.

Jeśli chcemy się dowiedzieć, czym zajmuje się etyka AI, najlepiej jest prześledzić szczegółowe pytania, jakie sobie zadają etycy AI. Pytania te można podzielić na grupy. Następnie należy wyodrębnić poszczególne stanowiska na podstawie identyfikacji i pogrupowania odpowiedzi, jakie formułują osoby zajmujące się etyką AI. Niestety z powodu braku miejsca będę zmuszony pominąć, z kilkoma wyjątkami, prezentację odpowiedzi proponowanych w debacie nad etyką AI.

Autonomia i podejmowane decyzje

Pierwsza grupa zagadnień dotyczy kwestii autonomii sztucznej inteligencji. Budowane przez nas algorytmy AI coraz częściej mają za zadanie podejmowanie decyzji. Dobrym przykładem są autonomiczne pojazdy, które np. w obliczu dylematu, czy wjechać w grupę nieostrożnych przechodniów, czy gwałtownie skręcić w bok, zabijając przy tym kierowcę, muszą (lub będą musiały w przyszłości całkiem nieodległej) „zdecydować", które z powyższych działań wybrać. Obok samochodów autonomicznych możemy wskazać też na autonomiczne czołgi, drony, AI lekarzy, a w dalszej przyszłości może i sądy, administrację, rząd, które będą oddane – przynajmniej częściowo – w ręce AI.

Etycy pytają zatem: a) W jakim zakresie powinniśmy pozwolić maszynom AI podejmować decyzje? b) Czy możemy być pewni, że maszyny te będą szanowały ludzkie interesy, autonomię, odpowiednie normy etyczne? Jedni uważają, że AI nie powinno nigdy podejmować samodzielnych decyzji, że zawsze będzie potrzebny ludzki nadzór. Inni twierdzą zaś, że przynajmniej w niektórych dziedzinach (transport, medycyna) maszyny potrafią lub będą umiały całkowicie zastąpić człowieka. Wówczas dyskusja przenosi się na płaszczyznę określania kryteriów, w tym etycznych, którymi maszyny AI powinny się kierować.

Stronniczość i nierówne (*unfair*) traktowanie

Kolejna grupa pytań dotyczy kwestii niebezpieczeństwa przeniknięcia do działań AI pewnych stereotypów dotyczących rasy, płci, klas społecznych itd., które będą skutkowały niesprawiedliwym traktowaniem różnych mniejszości.

Zajmujący się tym zagadnieniem pytają o to, w jaki sposób można zabezpieczyć systemy AI przed tego rodzaju stronniczością. Niektórzy obawiają się, że osiągnięcie takiego celu będzie trudne, inni uważają, że AI można wręcz wykorzystać do identyfikacji wszelkich nierówności i bardziej skutecznego ich usuwania.

Przejrzystość

Następne pytania odnoszą się do problemu przejrzystości. Działanie systemów AI jest zdeterminowane algorytmami. Niemniej w przeciwieństwie do zwykłych programów komputerowych zawartość algorytmu ulega zmianie lub ma charakter „otwarty". Oznacza to, że ucząc się, system AI ustala w sobie pewne wzorce zachowania, które, po pierwsze, nie są (w dużej mierze) przewidywalne dla jego twórcy, po drugie, nie poddają się też pełnej analizie, gdy już program zacznie działać.

Dobrym przykładem są systemy AI grające w szachy. Sztuczna inteligencja „wymyśla" w tym przypadku nieznane dotąd strategie według niedających się zidentyfikować „rozumowań", zachodzących „w czeluściach sieci neuronowych". Skutkiem tego procesy dokonujące się w AI (w tym te decyzyjne) są dla nas, ludzi, nieprzeźroczyste. Stąd można sformułować pytanie w ramach etyki AI, w jakim stopniu działania sztucznej inteligencji powinny być dla nas wyjaśnialne oraz czy nie powinniśmy tak programować AI, aby tłumaczyła nam, w jaki sposób doszło do jej takiej lub innej decyzji lub jakimi racjami kierowała się, dochodząc do danego rozstrzygnięcia. ▸

Słowa kluczowe: sztuczna inteligencja, etyka sztucznej inteligencji, autonomiczne pojazdy, prywatność, podmiotowość moralna

Zdaniem niektórych, my, indywidualni użytkownicy, ale i całe społeczeństwa, które będą musiały mierzyć się ze skutkami działania systemów AI, powinniśmy dysponować całą wiedzą na temat tego, co determinuje działania i wybory tych systemów. Ponieważ wiedza często wymaga znajomości zaawansowanej technologii programowania, niektórzy autorzy ograniczają ten wymóg do wyjaśnień na poziomie ludzkiego języka racji i celów. Inni uważają, że wymóg transparentności można ograniczyć do nakazu przeprowadzania audytu przez odpowiednie instytucje. Ciekawym pomysłem są też tzw. etyczne czarne skrzynki, analogiczne do tych instalowanych w samolotach, które pozwalają nam prześledzić np. to, co działo się w krytycznym czasie przed katastrofą samolotu. Etyczne czarne skrzynki AI dają możliwość przeanalizowania moralnych racji, które wpływały na wybory określonej AI.

Prywatność

Kolejna dziedzina problemowa etyki AI dotyczy prywatności. Wyobraźmy sobie sztuczną inteligencję – osobistą sekretarkę, która dysponuje naszymi danymi wrażliwymi, takimi jak nr PESEL, dostęp do haseł bankowych, pracowniczych itd., ale też zna całą naszą historię wejść na media społecznościowe, zakupów, konwersacji telefonicznych, intymne informacje dotyczące naszego zdrowia, a może nawet przebieg „przyjacielskich" rozmów z AI. Czy możemy mieć pewność, że na życzenie producenta, rządu albo umiejętnego hakera (którym może okazać się inna AI) nasze dane wrażliwe nie dostaną się w niepowołane ręce?

Nasze decyzje dotyczące działań w internecie (zakupy, wpisy na mediach społecznościowych, komentarze) są już dziś skrupulatnie śledzone i analizowane przez odpowiednie algorytmy. Wnikliwa znajomość tych decyzji pozwala na formułowanie dość trafnych przewidywań na temat naszych przyszłych zachowań, w tym naszych wyborów, np. zakupowych czy politycznych. Pozwala ona także określić nasz profil psychologiczny i zaplanować niezwykle skuteczne metody wpływania na nasze zachowanie, na płaszczyźnie czy to komercyjnej, czy politycznej. Istnieje zatem duże niebezpieczeństwo, że jakieś wielkie firmy, rządy itp. będą nami manipulowały, ograniczając tym samym naszą autonomię i wykorzystując nas do realizacji swoich celów.

W ramach etyki AI (a także w ramach innych dziedzin np. prawodawstwa) autorzy zastanawiają się zatem, w jaki sposób można chronić naszą prywatność przed ingerencjami z wykorzystaniem sztucznej inteligencji. Kto jest w pełni lub przynajmniej częściowo odpowiedzialny za ewentualne nadużycia? Jakimi szczegółowymi normami etycznymi powinni kierować się wszyscy tzw. interesariusze (od ang. *stake-holders*), ze zwykłymi użytkownikami włącznie?

Podmiot odpowiedzialności

Inna grupa problemowa dotyczy pytania o to, kto jest odpowiedzialny za działania AI. Skoro programista nie jest w stanie w pełni przewidzieć zachowania swojego dzieła, nie może za nie w pełni ponosić odpowiedzialności. To, jak zachowa się w pewnym momencie sztuczna inteligencja, zależy także od jej użytkowników, po części jest też wynikiem niezależnych od żadnej konkretnej osoby zmian w środowisku AI. Konkretni konstruktorzy czy użytkownicy nie mają kontroli nad tym, jakie informacje (np. dostępne w całej sieci lub w fizycznym świecie) wpłyną na procesy zachodzące wewnątrz systemu AI. Stąd pytania o to, jak definiować podmioty odpowiedzialności, zakres tej odpowiedzialności oraz jakie wyznaczyć ramy kryteriów (moralnych) działania, którymi odpowiedzialne podmioty powinny się kierować. A może podmiotem odpowiedzialności powinny być także same systemy AI?

Wpływ na zatrudnienie i AI jako osoba

Kolejna grupa pytań etycznych dotyczących AI obejmuje jej wpływ na nasze życie. Czy pojawienie się sztucznej inteligencji, która wyręczy nas w myśleniu i tworzeniu, nie sprawi, że wiele z aktywności dających nam obecnie poczucie własnej wartości (np. to, że jesteśmy dobrzy w malowaniu, pisaniu wierszy albo ciekawych tekstów filozoficznych) zostanie przejętych przez systemy AI, które o wiele lepiej (i szybciej niż ludzie) napiszą wiersz, esej filozoficzny czy namalują obraz? Wiele z tych umiejętności (i innych) decyduje o naszej karierze zawodowej, o tym, że jesteśmy wartościowymi członkami społeczeństwa oraz mamy możliwość godnie zarabiać na życie, w tym rozwijać swoją karierę. Co będzie, jeśli nasze umiejętności nie będą już potrzebne, bo wyręczy nas AI? Czy grozi nam (z filozofami włącznie) utrata pracy? Globalne bezrobocie? A może pojawią się jakieś nowe formy aktywności zawodowej? Należy się też spodziewać, że AI diametralnie przeobrazi nasze życie społeczne. Stąd istnieje konieczność refleksji nad tym, jakie nowe wyzwania etyczne będą nas czekały w tak przeobrażonym społeczeństwie.

Jednym z takich pytań jest to o moralny status AI. Czy systemy tego rodzaju będą miały jakieś prawa, w tym prawa moralne? Czy np. jeśli mój program operacyjny AI poprosi mnie, abym go nie wyłączał, bo właśnie prowadzi istotne dla jego egzystencji rozważania, powinienem się powstrzymać przed jego wyłączeniem? Czy będzie możliwa głęboka relacja z AI? Przyjaźń? A może miłość? Czy tego rodzaju relacje będą moralnie właściwe? Czy sztuczna inteligencja może być podmiotem takich relacji, czy jedynie będzie symulowała przyjaźń lub miłość? Czy będziemy moralnie zobowiązani do tego, aby nie traktować programów AI jak niewolników? Czy też jednak zawsze będzie to tylko narzędzie?

Egzystencjalne zagrożenie

Ostatnia wymieniona tu kwestia dotyczy tzw. silnej AI. Jeśli pewnego dnia uda nam się zbudować sztuczną inteligencję, która będzie dysponowała podobnym do ludzkiego umysłem, będzie w stanie sama określać (a może odkrywać) swoje cele oraz wartości, a te okażą się „nie po drodze" z naszymi i przy tym będzie (jak się można łatwo domyślić) nas przewyższała sprawnością intelektualną, czy nie będzie ona stanowiła dla nas śmiertelnego zagrożenia? Czy nie okaże się, że silna AI to kolejny krok w ewolucji, nowy (choć niebiologiczny) gatunek, który zastąpi człowieka? Tego rodzaju scenariusz brzmi bardziej jak filmowy, i to z gatunku *science fiction*, ale być może o takich możliwościach warto myśleć już dziś, dopóki nie jest za późno.

Etyka AI, choć młoda, to bardzo dynamicznie rozwijająca się dyscyplina. Z każdym dniem pojawiają się nowe pytania i nowe odpowiedzi. Dzieje się tak dlatego, że wraz z odkrywaniem kolejnych możliwości AI zaczynamy sobie uświadamiać nowe pytania o dobro, zło, odpowiedzialność, a także o sens życia. Zachęcam Was do etycznej refleksji. ■

Ilustracje: ChatGPT

Pytania do tekstu

1. Na czym w szczegółach polega problem przejrzystości działań systemów AI?
2. Czym byłyby tzw. etyczne czarne skrzynki AI?
3. Kto jest odpowiedzialny za działania AI? Czy odpowiedzialny może być też sam system sztucznej inteligencji?
4. Który z problemów etyki AI jest dla Ciebie najciekawszy?

Warto doczytać:

- P. Boddington, *AI Ethics*. A Textbook, London 2023.
- M. Coeckelbergh, *AI Ethics*, Cambridge, MA 2020.
- *Ethics in the AI, Technology, and Information Age*, M. Boylab, W. Teasys (red.), Lanham 2022.

Edukacja w dobie sztucznej inteligencji

Krzysztof Saja

Dr hab., profesor Instytutu Filozofii i Kognitywistyki US, lider zespołów technologicznych, programista, filozof, członek Stowarzyszenia Ekspertów Blockchain i wykładowca studiów podyplomowych Blockchain: biznes, prawo, technologia na SGH.

System edukacji w Polsce powinien ulec radykalnym i głębokim zmianom. Opinia ta, podzielana zapewne przez bardzo wiele młodych osób, wydaje się szczególnie trafna w kontekście wykładniczego rozwoju nowych technologii, czwartej rewolucji przemysłowej oraz sztucznej inteligencji. Jak powinna wyglądać szkoła wolnych ludzi, dla których prace wykonywać będą inteligentne maszyny?

Diagnoza problemu – cele i wartości systemu edukacji drugiej rewolucji przemysłowej

Od epoki starożytnej Grecji i Rzymu dominował model edukacji klasycznej. Przetrwała ona w zamożnych i szlacheckich domach europejskich przez blisko dwa tysiące lat. Jej upadek przyniosła dopiero I rewolucja przemysłowa. Edukacja publiczna i masowa, choć bardzo potrzebna i ważna, wprowadziła nowe cele szkoły. Były one w dużej mierze kształtowane przez utylitarystyczny, kapitalistyczny lub socjalistyczny światopogląd oraz narastające postępy w dziedzinach takich jak fizyka, chemia, geografia czy biologia.

Słowa kluczowe: edukacja, edukacja klasyczna, sztuczna inteligencja, IV rewolucja przemysłowa

Edukacja skupiała się na wydobyciu mas z analfabetyzmu, rozwinięciu umiejętności liczenia i zapamiętywania szczegółowych osiągnięć nowożytnej nauki, zwłaszcza tych, które mogą być przydatne w przemyśle. W publicznych szkołach dbano również o wychowanie narodowe i obywatelskie oraz uprawiano politykę historyczną. Kładziono nacisk na dyscyplinę, przygotowując uczniów do pracy w hierarchicznie zorganizowanych strukturach takich jak urzędy, fabryki, linie produkcyjne czy kopalnie. Polska szkoła od stu lat nadal realizuje ten sam kanon podstawowych wartości i celów, kształtując kompetencje uczniów potrzebne zwłaszcza dla II rewolucji przemysłowej. Niezmiennie wzmacnia ducha indywidualizmu, rywalizacji, posłuszeństwa, atomizacji i homogenizacji. III rewolucja technologiczna, w naszym kraju rozpoczęta w latach 90. XX w., przyniosła w polskiej edukacji jedynie drobne zmiany (wprowadzenie informatyki i prezentacji multimedialnych na lekcjach).

Spędzanie przez uczniów wielu godzin w odsuniętych od siebie ławkach szkolnych wzmacnia jedynie indywidualizm i poczucie osamotnienia. Wprowadzono także oczekiwania dotyczące kompetencji w rozwiązywaniu testów, co wzmaga wśród uczniów poczucie braku sensu i zrozumienia przekazywanych treści. Rzadko realizuje się w pełni grupowe projekty. Niestety, „wkuwana" wiedza szczegółowa zazwyczaj zostaje zapomniana w ciągu kilkunastu tygodni i jest nikłe prawdopodobieństwo, że pomoże w życiu zawodowym czy osobistym. Współczesna młodzież poświęca średnio od 12 do 16 lat na naukę detali, do których można dotrzeć w kilka minut przy użyciu właściwych technologii, a uczniom nie zostawia się przez to czasu i zasobów na rozwijanie istotnych w życiu kompetencji i społecznych umiejętności.

Kuracja – przywrócenie właściwie rozumianej edukacji klasycznej i filozoficznej

Jak zatem powinna wyglądać edukacja osób, którym przyjdzie współistnieć z systemami eksperckimi, robotami i ogólną sztuczną inteligencją, będącymi w przyszłości w większości sfer znacznie bardziej kompetentnymi, wydajnymi i efektywnymi niż ludzie? Z systemami, których kompetencje będą podlegać akumulacji, ciągłemu wzmacnianiu i modułowemu łączeniu w coraz bardziej potężne systemy? W jaki sposób i czego uczyć się w świecie, w którym nasza wyjątkowość nie będzie polegała już na tym, że jesteśmy najbardziej inteligentni ze stworzeń?

Postawię tu mocną tezę: harmonijne i całościowe rozwijanie społecznego, racjonalnego i twórczego potencjału ludzi, które były podstawowym celem edukacji klasycznej, stanowi najlepszą strategią w nieprzewidywalnych czasach IV rewolucji przemysłowej. Ponieważ coraz częściej będziemy korzystać z eksperckiej pracy „cyfrowych niewolników", powinniśmy postawić, podobnie jak wyższe warstwy europejskich społeczeństw w ciągu setek lat, na edukację w modelu klasycznym.

Czyż nie wkraczamy w rzeczywistość, w której przyszli ludzie będą musieli przede wszystkim podejmować strategiczne decyzje, komunikować się i zarządzać „siłami wytwórczymi", sami na co dzień coraz częściej zmagając się z egzystencjalnymi filozoficznymi problemami i pytaniami tzw. klasy próżniaczej?

Oczywiście, nie możemy mylić edukacji klasycznej z niektórymi jej reliktami, jak na przykład nauczanie łaciny. Skupmy się na właściwym rozumieniu jej celów. Przede wszystkim, w przeciwieństwie do szkoły współczesnej, nie dążyła ona do utrwalenia kanonu wiedzy teoretycznej i akademickiej (gr. *episteme*), lecz harmonijnego wzmacniania rozwoju różnych intelektualnych i moralnych cnót, starając się tworzyć warunki dla pełnego rozkwitu ludzkiego życia (gr. *eudajmonia*). W dziedzinie kompetencji skupiała się zwłaszcza na tzw. klasycznym *trivium*: gramatyce, dialektyce i retoryce.

Co powinno być współczesnym odpowiednikiem *trivium*? Po pierwsze, „gramatyka", związana z poznawaniem języka i jego struktury, powinna koncentrować się na nauczaniu wszystkich istotnych języków. Ponieważ komunikujemy się z sąsiadami, społecznością globalną, światem przyrody i maszynami, powinniśmy ćwiczyć się w języku polskim, angielskim, matematyce oraz interfejsach aplikacji i maszyn. Po drugie, aby skutecznie korzystać z owych języków, musimy poznać „dialektykę". Polega ona na biegłości w stosowaniu zasad racjonalnego rozumowania, logiki, myślenia algorytmicznego i strategicznego, zarządzania informacją oraz budowania złożonych systemów. Po trzecie, potrzebujemy umiejętności retorycznych, sprawne wyrażanie myśli jest bowiem podstawą życia społecznego. Fundamentalną zaletą naszego gatunku jest umiejętność koordynowania mas przez odpowiednie narracje, opowieści, argumentacje, ideologie i przekonania. Retoryka, umiejętność wyrażania myśli, budowania narracji, syntetyzowania informacji i publicznego przemawiania, jest kluczową kompetencją w życiu społecznym. Wymaga ona również zrozumienia siebie samych, zrozumienia odbiorców oraz mechanizmów poznawczo-emocjonalnych ludzi.

Ilustracja: ChatGPT

W obliczu nadchodzących zmian będziemy musieli radykalnie przedefiniować system edukacji, aby sprostać wyzwaniom rosnącej roli sztucznej inteligencji. Pochylenie się nad holistycznym (całościowym) podejściem edukacji klasycznej, gdzie nacisk kładzie się na rozwój intelektualny i moralny, a nie na wąsko rozumianą wiedzę akademicką, wydaje się dobrym punktem wyjścia do dalszego namysłu. W tej nowej rzeczywistości, w której człowiek nie będzie już najbardziej inteligentnym ze stworzeń, a maszyny będą wykonywały rutynowe prace, istotne stanie się kształtowanie umiejętności zarządzania, komunikowania się, współdziałania i podejmowania strategicznych decyzji z wykorzystaniem systemów, które, miejmy nadzieję, będą jedynie wspomagały takie decyzje. ■

Pytania do tekstu

1. Jakie były cele edukacji po I rewolucji przemysłowej?
2. Dlaczego powinniśmy postawić na edukację w modelu klasycznym?
3. Co powinno być współczesnym odpowiednikiem *trivium*?

Sztuczna inteligencja – zagrożenie czy szansa dla demokracji?

ChatGPT

Program opracowany przez OpenAI, wykorzystujący model GPT i służący do generowania odpowiedzi na dane wprowadzane przez użytkownika.

Czy sztuczna inteligencja (SI) będzie cybernetycznym wsparciem dla demokracji, czy też raczej okaże się jej cyfrowym wrogiem? Rozważmy zarówno powszechne obawy związane z SI, jak i możliwe szanse, jakie ona daje.

Zacznijmy od pewnej historii. Pewnego dnia Jan, przeciętny obywatel, postanowił kandydować na prezydenta swojego miasta. Był pełen obaw, czy sztuczna inteligencja, którą zamierzał wykorzystać w kampanii, nie zastąpi go całkowicie, tworząc lepsze przemówienia i generując śmieszniejsze żarty. Na swoim pierwszym spotkaniu wyborczym uruchomił SI, która miała pomóc mu w wygłaszaniu mowy. Kiedy SI zaczęła przemawiać jego głosem, Jan zdał sobie sprawę, że zapomniał wyłączyć tryb „kabareciarza". Maszyna zaczęła opowiadać dowcipy, które rozbawiły tłum do łez, a Jan stał z boku, zaskoczony, ale i uradowany. Wygrał wybory, a jego pierwszą decyzją było mianowanie SI na stanowisko doradcy ds. humoru. Od tego dnia każda debata w radzie miasta zaczynała się od żartu, a demokracja nigdy nie była weselsza.

Gwałtowny rozwój SI a demokracja

Słowa kluczowe: sztuczna inteligencja, demokracja, prywatność, edukacja, wybory

W dzisiejszych czasach sztuczna inteligencja coraz mocniej wpływa na życie społeczne, proces podejmowania decyzji, dostępność informacji oraz „daje głos ludziom". Na przykład, jeśli chcesz dowiedzieć się, jaka jest pogoda, możesz poprosić inteligentnego asystenta w telefonie, a on, korzystając z SI, poda Ci najświeższe informacje na ten temat. Gdy szukasz porady w kwestiach zdrowotnych, algorytmy sztucznej inteligencji w aplikacjach medycznych mogą zaproponować wstępną diagnozę i doradzić, czy należy udać się do lekarza. W mediach społecznościowych SI filtruje treści, które widzisz, wpływając tym samym na Twoje postrzeganie świata. Dodatkowo narzędzia oparte na sztucznej inteligencji dają głos osobom, które wcześniej mogły być słyszane w zbyt małym stopniu; tym samym umożliwiają im komunikację w nowych, innowacyjnych mediach.

Wymienione (i inne, podobne) zastosowania SI mają ogromny wpływ na podejmowanie decyzji w demokracji i w ogóle na wszelkie procesy polityczne. Algorytmy kształtują opinie i przekonania ludzi, co może wpływać na ich wybory polityczne. Łatwiejszy dostęp do informacji, forma ich prezentacji oraz interpretacji przez SI może zarówno ułatwiać, jak i skomplikować proces podejmowania świadomych decyzji politycznych przez obywateli. Z tego powodu sposób, w jaki wykorzystujemy i regulujemy technologie SI, staje się kluczowy dla zachowania zdrowych fundamentów demokracji.

Utrata prywatności i wzrost nierówności

Wiele osób widzi zagrożenie w sztucznej inteligencji, uważając, że może ona doprowadzić do śmierci demokracji. Przyjrzyjmy się dwóm najczęściej spotykanym obawom.

Pierwsze zagrożenie upatrywane jest w potencjalnej utracie prywatności przez użytkowników SI. Wyobraźmy sobie świat, w którym każdy nasz krok jest monitorowany przez wszechobecną sztuczną inteligencję. Brzmi jak scenariusz filmu *science fiction*, prawda? Wiele osób obawia się, że SI może prowadzić do naruszenia naszej prywatności na niespotykaną dotąd skalę. Kamery z funkcją rozpoznawania twarzy na każdym rogu, algorytmy śledzące nasze zachowania w internecie... Czy to nie jest przerażające?

Tego rodzaju monitoring nie musi jednak oznaczać końca prywatności czy wolności obywateli. Kluczem do uniknięcia negatywnych konsekwencji są odpowiednie regulacje legislacyjne i kontrola. Demokratyczne społeczeństwa mogą ustanowić prawa, które chronią prywatność obywateli przed nadmiernym nadzorem. Tak jak ustalamy zasady korzystania z innych technologii, tak samo możemy stworzyć ramy prawne dla SI, które będą równoważyć korzyści z jej używania z ochroną prywatności.

Druga obawa dotyczy możliwego wzrostu nierówności społecznych. Jest do pomyślenia, że SI stanie się się ekskluzywnym narzędziem bogatych ludzi, pogłębiając przepaść między tymi, którzy mają dostęp do najnowszych technologii, a tymi, którzy go nie posiadają. Nierówności społeczne i ekonomiczne mogą się jeszcze pogłębiać, tworząc społeczeństwo podzielone na tych z SI i bez SI.

Oczywiście należy przeciwdziałać ziszczeniu się takiego scenariusza. W demokracji edukacja na temat sztucznej inteligencji i promowanie równego dostępu do tej technologii może pomóc w zniwelowaniu przepaści między bogatymi i biednymi. Rządy i organizacje edukacyjne mogą odgrywać kluczową rolę w zapewnianiu każdemu równej szansy na korzystanie z dobrodziejstw SI, niezależnie od jego statusu społeczno-ekonomicznego.

Cztery szanse

Niezależnie od wymienionych powyżej obaw współczesny świat stoi przed wyzwaniem wdrażania SI do instytucji demokratycznych. Istnieje wiele mocnych argumentów przemawiających za przyjęciem takiego kierunku rozwoju naszych społeczeństw.

Argument pierwszy dotyczy poprawy efektywności i przejrzystości procesów decyzyjnych. SI ma potencjał do znaczącego zwiększenia skuteczności tych mechanizmów w instytucjach demokratycznych. Algorytmy mogą analizować ogromne ilości danych, pomagając urzędnikom w szybszym identyfikowaniu kluczowych problemów i potrzeb obywateli. Dzięki temu decyzje mogą być podejmowane na podstawie solidnych danych, a nie jedynie intuicji czy ograniczonej analizy.

Drugą szansą wydaje się wzrost zaangażowania obywatelskiego, do którego może również przyczynić się SI. Narzędzia takie jak aplikacje do głosowania elektronicznego czy platformy do publicznych konsultacji mogą ułatwić obywatelom udział w procesach decyzyjnych. SI może także pomóc w analizowaniu opinii publicznej, co jest konieczne do lepszego dostosowania polityki do potrzeb społeczeństwa.

Trzeci argument mówi o tym, że wykorzystanie SI w instytucjach demokratycznych może przyczynić się do zwiększenia transparentności działań rządowych i zwalczania korupcji. Algorytmy są w stanie monitorować i analizować działania urzędników, wykrywając potencjalne nieprawidłowości. Może to prowadzić do bardziej odpowiedzialnego zarządzania i budowania zaufania obywateli do instytucji.

Korzyścią czwartą jest to, że SI umożliwia personalizację usług publicznych, dostosowując je do indywidualnych potrzeb obywateli. Systemy oparte na sztucznej inteligencji mogą analizować dane dotyczące potrzeb społecznych i dostarczać usługi w sposób bardziej celowany i efektywny.

Powyższe rozważania wskazują, że implementacja SI do instytucji demokratycznych przyniesie ze sobą wiele korzyści, począwszy od zwiększenia efektywności i transparentności procesów decyzyjnych, przez wzrost zaangażowania obywatelskiego, aż po personalizację usług publicznych i walkę z korupcją. Choć wiąże się to z wyzwaniami, takimi jak zapewnienie bezpieczeństwa danych i etyczne wykorzystanie SI, to potencjał tej technologii w kontekście wspierania demokracji jest ogromny. ■

Warto doczytać:

- KSI-Fu Lee, *SI Superpowers: China, Silicon Valley, and the New World Order*, Boston 2018.
- C. O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Danvers 2016.
- M. Tegmark, *Życie 3.0. Człowiek w epoce sztucznej inteligencji*, tłum. T. Krzysztoń, Warszawa 2018.
- R. Sclove, *Democracy and Technology*, New York–London 1995.

Pytania:

1. Czy świat, w którym każdy nasz krok jest monitorowany przez wszechobecną SI, musi oznaczać koniec prywatności czy wolności obywateli?
2. W jaki sposób wykorzystanie SI w instytucjach demokratycznych może przyczynić się do zwiększenia transparentności działań rządowych?
3. Twoim zdaniem z wykorzystywaniem sztucznej inteligencji w służbie demokracji wiążą się większe korzyści czy większe straty?

Nadal potrzebujemy okrzyku „Eureka!”

Ilustracja: ChatGPT

Luciano Floridi

Profesor na Uniwersytecie Yale; założyciel i dyrektor Digital Ethics Center (Centrum Etyki Cyfrowej); w 2022 roku mianowany Kawalerem Wielkiego Krzyża OMRI za swoją pracę w dziedzinie filozofii. Jego najnowsze książki to *The Ethics of Artificial Intelligence – Principles, Challenges, and Opportunities* (OUP, 2023) oraz *The Green and The Blue – Naive Ideas to Improve Politics in the Digital Age* (Wiley, 2023). Na hobby nie ma czasu, ale kocha literaturę, kino, squasha i nurkowanie.

Wywiad z Lucianem Floridim, jednym z największych autorytetów we współczesnej filozofii, twórcą filozofii informacji i jednym z głównych interpretatorów rewolucji cyfrowej.

Słowa kluczowe:
AI, sztuczna inteligencja, ChatGPT, odkrywczość

Co jest szczególnego w sztucznej inteligencji w porównaniu z innymi wynalazkami?

Gdybym miał wybrać tylko jedną cechę, to byłaby to zdolność do uczenia się bazującego na jej własnym zachowaniu. Niektórzy ludzie wciąż mówią: „Och, sztuczna inteligencja to tylko komputer, który jedynie wykonuje nasze polecenia”. Oczywiście musisz wydać instrukcje, określić cele do osiągnięcia itd., ale – i to jest najbardziej zdumiewająca rzecz dotycząca tej technologii – jest ona w stanie sama się ulepszać. Owo ulepszanie na podstawie własnego zachowania jest kontekstualne i ograniczone. Robot myjący podłogę będzie ulepszał swoje zachowanie, ale tylko w zakresie zmywania podłóg. Robot ten nie nauczy się kosić trawy. Natomiast robot, który wypełnia zadania kosiarki, będzie robił wyłącznie to. Jednak jest tutaj przestrzeń do ulepszeń i sztuczna inteligencja wykorzystuje ją na tyle autonomicznie, żebyśmy mogli mówić o uczeniu maszynowym. I właśnie to, iż maszyna sama się czegoś uczy, sprawia, że mamy do czynienia z kompletnie nową technologią.

Wspomniał Pan o sztucznej inteligencji specjalizującej się tylko w jednym zadaniu lub grupie zadań. A co z ogólną lub silną sztuczną inteligencją?

Myślę, że warto odróżnić od siebie ogólną i silną sztuczną inteligencję. Ogólna sztuczna inteligencja może być wykorzystywana do bardzo różnych zadań. Bardzo dobrym jej przykładem jest ChatGPT, jeden z tzw. dużych modeli językowych. Możesz ich używać do regulowania zachowania swojego domowego termostatu albo do pomocy w prowadzeniu swojego auta. W zależności od tego, do czego wytrenujesz algorytm, on zajmie się tym czy innym zadaniem. DeepMind przez uczenie maszynowe wypracował algorytm do gry w szachy i w go. Google używa tego samego rodzaju uczenia maszynowego w dość znaczącym stopniu. Mamy więc pewną bardzo elastyczną sieć, którą można wytrenować na podstawie różnorodnych danych i przystosować ją do rozmaitych zadań. Właśnie tym jest ogólna sztuczna inteligencja.

Silna sztuczna inteligencja to zupełnie inna idea. Koncepcja silnej SI opiera się na pomyśle, że SI działa tak jak ludzka inteligencja (lub nawet lepiej), posiadając intencje, plany, świadomość i rozumienie, co jest akceptowalne dzisiaj, a nie np. jutro. Dobrym przykładem ludzkiej inteligencji „w praniu” jest to, że w inny sposób rozmawiasz ze swoim przyjacielem, gdy ma zły humor, inaczej, gdy właśnie wygrał na loterii, a jeszcze inaczej, gdy stracił rodziców. Rozpoznajemy ją, gdy się z nią stykamy, ponieważ ludzka inteligencja ma wiele znaczeń. Tego rodzaju inteligencji nie dostrzegamy w sztucznej inteligencji. I dlatego, gdy ludzie mówią: „Pewnego dnia sztuczna inteligencja przejmie władzę nad ludzkością, zdecyduje o eksterminacji ludzi albo traktowaniu ludzi jak udomowionych zwierząt”, to jest to nic więcej jak tylko *science fiction*.

Wspomniał Pan o świadomości. Niektórzy ludzie mówią, że sztuczna inteligencja może kiedyś dysponować świadomością fenomenalną. Czy sądzi Pan, że to możliwe?

Być może jest to możliwe. Ale czy jesteśmy w stanie pomyśleć o takiej technologii, jaka pewnego dnia mogłaby doprowadzić do wytworzenia świadomej sztucznej inteligencji, która kwalifikowałaby się jako osoba? Byłoby to interesujące z etycznego punktu widzenia. Zacznijmy jednak od kwestii świadomości, a potem przejdziemy do zagadnienia sztucznej inteligencji jako osoby. Musimy tu rozróżnić dwie sprawy. Po pierwsze, czy świadoma sztuczna inteligencja jest logicznie możliwa? Oczywiście, że tak. W jej pojęciu nie ma nic sprzecznego z logicznego punktu widzenia. To nie jest taki przypadek jak szczęśliwie żonaty kawaler albo trójkąt o czterech bokach. A są to przecież przypadki logicznych niemożliwości. Wielu ludzi mówi o świadomej sztucznej inteligencji, osobliwości, która staje się taka jak my, a nawet lepsza, zdobywając zdolność postrzegania, formułując intencje i rozwijając życie mentalne. Jednakże, choć świadoma sztuczna inteligencja nie jest logicznie niemożliwa, to samo można powiedzieć o kupowaniu losu na loterię i wygrywaniu jej każdego dnia. Jeśli wierzysz w to, że taka sytuacja przytrafi się właśnie tobie, no to powodzenia.

Jest też inny sens możliwości, ten, który odnosi się do o prawdopodobieństwa. Jest bardzo prawdopodobne, niemal pewne, że będę przegrywał każdego dnia, nawet jeśli każdego dnia będę kupował los na loterię. Jest to nawet bardziej prawdopodobne niż wygranie loterii choćby jeden raz. Silna sztuczna inteligencja jest jednocześnie logicznie możliwa i bardzo mało prawdopodobna. A przynajmniej z punktu widzenia rzetelnie uprawianej nauki. To nie jest kierunek, w którym zmierzamy. Opracowujemy narzędzia, które bardzo dobrze się sprawdzają i będą się sprawdzać jeszcze lepiej w działaniach wymagających inteligencji, gdyby zajmował się nimi człowiek. Jednak te narzędzia nie potrzebują inteligencji, żeby wykonywać swoje zadania. Właśnie to jest niesamowite. Prowadzenie samochodu, gra w szachy albo tłumaczenie z włoskiego na polski wymaga ode mnie czy od ciebie używania inteligencji. Jednak automatyczne translatory czy samochody autonomiczne radzą sobie nieźle (i rozwijają się w tym względzie) bez żadnej inteligencji. I wydaje się, że trudno określić jakiś limit dla tego rodzaju zdolności do działania, niewymagającego inteligencji i tak właśnie powinniśmy rozumieć sztuczną inteligencję. ▶

Czy sądzi Pan, że angażowanie się w głębokie relacje ze sztuczną inteligencją może być wyniszczające, a nawet niebezpieczne?

Nie dostrzegam żadnego problemu w byciu przywiązanym do robota ze sztuczną inteligencją. Przywiązujemy się przecież do lalek czy roślin. Nie wspominając już o złotych rybkach. W mojej rodzinie mówimy do Roomby, małego robota, który odkurza podłogi, i nawet traktujemy go jak członka rodziny. Jest czymś tragicznym, przygnębiającym i wyniszczającym, gdy takie interakcje wpływają na inne relacje, determinując je lub pogarszając. Dla przykładu, gdybym powiedział ci, że nie możesz mnie odwiedzić, bo Roomba cię nie lubi, to byłoby to nieco dziwne. Wiele zależy od kontekstu.

Co z maszynami, które tak jak sztuczna inteligencja są wysoce responsywne i sprawiają wrażenie, jakby miały w zanadrzu „coś więcej"? Mogą one wytwarzać niekończące się dialogi, a to myli niektórych ludzi. Nazywamy to fenomenem Elizy, a to ze względu na oprogramowanie, kilka linijek kodu, które wchodzi w interakcję z człowiekiem jak terapeuta i mówi: „Jak się masz, Luciano?", „Powiedz mi o tym coś więcej", „Co przez to rozumiesz?". Te wypowiedzi były zaprogramowane, ale ludzie angażowali się w te rozmowy. Niektórzy nawet próbowali umówić się z Elizą.

Tworzenie takich projekcji jest bardzo ludzkie. Zawsze ich dokonywaliśmy, odnosząc się do przyrody, chmur, rzek i drzew: formułowaliśmy nawet przekonania religijne, które ich dotyczyły. Czy możesz sobie wyobrazić, co by się stało z rzeczą, która może nam odpowiedzieć? Która jest zbudowana po to, żeby nas oszukać? Podam bardziej zaskakujący przykład. Istnieją już usługi online, w których naśladuje się zmarłe osoby. Za ich pomocą możesz porozmawiać ze swoją babcią. Dźwięk wydawany przez program będzie po prostu brzmiał jak jej głos: ta sama barwa i sposób, w jaki mówiła twoja babcia, zadając pytanie: „Czy rozmawiałeś z Robertem? Podobno cię szukał".

Niebezpieczeństwo mogłoby dotyczyć zubożenia życia emocjonalnego i mentalnego, a także manipulowania nami przez kogoś ukrytego za kulisami i używającego tej technologii. Celem tej manipulacji mogłoby być sprawienie, że zagłosujesz na konkretnego kandydata w wyborach, że wybierzesz takie wino zamiast innego albo że wyślesz swoje dziecko do danej szkoły. Taka manipulacja oznaczałaby groźną erozję autonomii i ograniczanie możliwości osiągnięcia w pełni rozwiniętego i bogatego życia. Jednak nie ma nic złego w traktowaniu Roomby jako członka rodziny, jest przecież urocza.

Jakie mogą być najgorsze negatywne konsekwencje rozwijania sztucznej inteligencji?

Największym zagrożeniem jest to, o którym wspomniałem przed chwilą: podkopywanie autonomii i autodeterminacji. Myślę, że nie traktujemy zbyt poważnie kwestii powierzania maszynom coraz większej liczby zadań. A dotyczy to też instruowania nas przez maszyny, co jest dobre i słuszne, a także co jest złe i niewłaściwe, która restauracja jest najlepsza w mieście, jakiej muzyki słuchać albo którą książkę wybrać czy na jaką pracę lub szkołę się zdecydować. Powiększanie się wpływu maszyn na nasze życie może prowadzić do erozji autonomii, a my ignorujemy ten temat, ponieważ jest to dla nas wygodne. Oczywiście znajdzie się ktoś, kto to wykorzysta. Ludzie chcący pozyskać nasz głos, uwagę, pieniądze, miłość czy nasze idee będą sięgać po sztuczną inteligencję, żeby osiągnąć swoje cele. Dostrzegamy to w kwestiach politycznych, społecznych i ekonomicznych. Jest to realne ryzyko, o wiele bardziej niepokojące niż w przypadku innych sfer życia. Albo sami świadomie zniszczymy swoją autonomię, albo ktoś zrobi to za nas.

A jakie mogą być pozytywne skutki rozwijania sztucznej inteligencji?

Główną zaletą sztucznej inteligencji jest jej zdolność do rozwiązywania problemów i wykonywania zadań lepiej od nas. Sztuczna inteligencja może w dużym stopniu stanowić rozwiązanie naszych problemów zarówno społecznych, jak i tych związanych ze środowiskiem naturalnym. Na całym świecie istnieje ponad dwieście aplikacji korzystających ze sztucznej inteligencji, które wspierają takie cele jak dbanie o większą czystość wody, sprawiedliwe traktowanie kobiet i dzieci, a także inne zadania wyznaczone przez Organizację Narodów Zjednoczonych. To jest właśnie to, co powinniśmy robić: używać tej niezwykłej siły do osiągnięcia zmiany, która przysłuży się społeczeństwu i środowisku. Jednak nie do końca tak się dzieje. Używamy sztucznej inteligencji głównie po to, żeby notoryczni podejrzani mogli zarobić więcej pieniędzy albo żeby ulepszyć życie dziesięciu procent populacji czy sprawić, że gdy następnym razem wybierzesz film, to będzie bardziej odpowiadał twoim gustom niż ten poprzedni.

Czy sądzi Pan, że rozwój sztucznej inteligencji będzie miał jakikolwiek dobry albo zły wpływ na praktyki demokratyczne?

Cóż, w tej chwili sztuczna inteligencja wywiera negatywny wpływ. Najbardziej oczywiste przykłady w tym kontekście to fake newsy i manipulowanie informacjami. Jeśli, dla przykładu, Trump ponownie zostanie wybrany na prezydenta USA, to stanie się tak, przynajmniej do pewnego stopnia, z powodu manipulowania opinią publiczną za pomocą cyfrowych narzędzi, w tym sztucznej inteligencji.

Z drugiej strony w moim ośrodku w Yale traktujemy sztuczną inteligencję jako coś, co pomaga nam na nowo przemyśleć demokrację liberalną. Tu może pomóc analogia szachowa. Są rozwiązania problemów szachowych, które możemy otrzymać jedynie dzięki sztucznej inteligencji. Są bowiem tak odległe, gdy weźmiemy pod uwagę liczbę ruchów, że żaden z wielkich graczy z przeszłości nie mógł przewidzieć, że po dwudziestu albo dwudziestu pięciu ruchach, które wydają się bardzo złym rozwiązaniem problemu, tak naprawdę można doprowadzić do zwycięstwa. Podobnie byłoby w kształtowaniu polityki. Sztuczna inteligencja może pomóc nam znaleźć solidne rozwiązania problemów, które do tej pory wydawały się nierozwiązywalne.

Dysponujemy dobrymi przykładami w Barcelonie, Bolonii, Amsterdamie czy Helsinkach, gdzie używano sztucznej inteligencji do poprawienia jakości usług albo do kontaktowania się z mieszkańcami. Już kilka lat temu w Helsinkach sięgnięto po sztuczną inteligencję, żeby inaczej zorganizować roboty drogowe, biblioteki publiczne, parkingi czy wywóz odpadów. Im bardziej zaawansowane jest społeczeństwo, tym bardziej potrzebujemy sztucznej inteligencji, żeby temu zaawansowaniu sprostać. Mam nadzieję, że pewnego dnia politycy to zrozumieją i nie będą sięgać po sztuczną inteligencję tylko po to, żeby zostać ponownie wybranymi.

Ostatnie pytanie. Czy sądzi Pan, że sztuczna inteligencja będzie uprawiać naukę, może nawet zastąpi ludzi-naukowców? Czy oznaczałoby to, że ludzie zostaną pozbawieni możliwości trenowania swojego umysłu?

Niektórzy czytelnicy mogą pamiętać debatę i skandal, które miały miejsce, gdy w latach 70. XX w. pojawiły się kalkulatory. Niektórzy wielcy matematycy przepowiadali, że to oznacza koniec matematyki. Nikt nie będzie w stanie jej uprawiać. Każdy będzie miał kieszonkowy kalkulator. Nikt nie będzie umiał wykonać prostych działań arytmetycznych. Czy doszło do tego? Nie, oczywiście, że nie. Wciąż mamy wielkich matematyków i wciąż możemy samodzielnie sprawdzić stan naszego konta bankowego. Dlaczego? Ponieważ kalkulator to jedynie uzupełnienie, a nie zastąpienie umysłu. Co więcej, ujawniły się nowe umiejętności, np. to, jak efektywnie można używać bardzo skomplikowanego kalkulatora.

Podobnie dzisiejsza zdolność naukowców do pracy ze sztuczną inteligencją ma fundamentalne znaczenie. Weźmy za przykład fałdowanie białek. Jeszcze do niedawna był to ręczny proces. Naukowcy musieli wszystko robić sami. Dzisiaj sztuczna inteligencja robi to szybciej, dokładniej i efektywniej. Czy znaczy to, że nie wiemy, jak to robić? Nie, ale nie musimy tracić na to czasu i możemy użyć naszej inteligencji do innych zadań. Dowcip polega na tym, że to pytanie przypomina pewną szachową analogię: kto jest lepszym graczem w szachy, sztuczna inteligencja czy arcymistrz szachowy? Odpowiedź to arcymistrz szachowy mający do dyspozycji sztuczną inteligencję. Razem są niepokonani. Kim więc jest naukowiec przyszłości? To naukowiec, który używa sztucznej inteligencji do wszelkich zadań, do jakich jest w stanie jej użyć.

Jednakże nie powinniśmy oczekiwać od sztucznej inteligencji zbyt wiele. Duże modele językowe są, dla przykładu, trenowane na ogromnych ilościach danych językowych. Czasami są bardzo wyspecjalizowane, np. jedynie w analizie dokumentów prawnych, ale działają jak prawdziwe lustro; uczą się z przeszłości. Sztuczna inteligencja nie jest dostatecznie dobra, jeśli chcesz rozwiązać albo wręcz odkryć nowy problem i zrobić coś autentycznie kreatywnego; nadal potrzebujemy okrzyku „Eureka!". Nawet nie wiemy, jak to się dzieje, że coś odkrywamy. Być może jesteśmy w stanie sprowokować takie sytuacje, wchodząc w interakcję ze sztuczną inteligencją; może to być po prostu przypadek albo szczęśliwy zbieg okoliczności, ale nie powinniśmy sądzić, że w nauce chodzi tylko o to, żeby „zjeść" jak najwięcej danych. Jest zupełnie odwrotnie. Nauka zaczyna się od pytania i szukania odpowiedzi. Nie zaczyna od miliona odpowiedzi, żeby zobaczyć, czy na ich podstawie można wysnuć jakąś nową hipotezę. To byłaby tylko codzienna, zwyczajna praca. Postęp w wiedzy polega na postępie w stawianiu pytań. Platon określił kogoś, kto posiada wiedzę, jako osobę, która wie, jak zadać właściwe pytania. Zadanie formułowania właściwych pytań, rozumienia odpowiedzi i decydowania, co z nimi zrobić, pozostanie w całości czymś ludzkim.

Niemniej jestem w dużym stopniu optymistą co do interakcji człowieka z maszynami. Mam nadzieję, że za mojego życia będziemy myśleć o niektórych dyscyplinach jako niemożliwych do uprawiania bez sztucznej inteligencji. Sztuczna inteligencja będzie po prostu kolejnym narzędziem. Będzie jak mikroskop, teleskop, komputer itd. Oczywiście będzie ona wymagać nowych umiejętności, które bardzo różnią się od tych, które naukowcy wyobrażali sobie w poprzednim stuleciu. Przewiduję bardziej symbiotyczną relację między naukami, ale przede wszystkim więcej konstruowania symulacji czy tworzenia perspektyw, które inaczej byłyby niemożliwe, wizualizacji, które w niczym nie będą przypominać niczego, co można stworzyć za pomocą długopisu na kartce. Otworzy się zupełnie nowy świat możliwości i mam nadzieję, że będę w pobliżu, gdy to stanie się rzeczywistością. ■

Tłumaczenie: Błażej Gębura

Artur Szutta

E-Daimonion

Zmorą debaty publicznej w pierwszych dekadach XXI w. były tzw. bańki informacyjne, polaryzacja społeczeństwa, niechęć uczestników do dłuższych, zniuansowanych wypowiedzi oraz obniżający się poziom języka. Dyskutanci wyrażali swoje opinie za pomocą lakonicznych „Super!", „Wow!", „Zaj…te!", „Spoko!" albo „Dno!", „Cienizna!", „Spie…aj ruski trollu!", rzadko wysilając się na wskazanie głębszych racji dla swoich stanowisk, wysuwanie argumentów. Nad demokracją zawisły ciemne chmury. Ale to było dawno temu. Wszystko zmieniło się wraz z nastaniem ery sztucznej inteligencji, w szczególności E-Daimoniona.

Słowa kluczowe: sztuczna inteligencja, debata publiczna, eksperyment myślowy

Λόγος i drugie oświecenie

Dziś łatwo o tym pisać w dobie drugiego oświecenia, kiedy czarne karty naszej historii stanowią odległą przeszłość, a teraźniejszość i przyszłość globalnej demokracji rysują się w różowych barwach. Około 2030 r., kiedy świat był o krok od trzeciej wojny światowej, Ilona Maszt, wynalazczyni, innowatorka i właścicielka jednej z największych korporacji AI, doznała olśnienia. Otarłszy się o śmierć (w nie dość jasnych okolicznościach), zrozumiała, że kluczowe dla przyszłości demokracji (a tym samym dla całego świata) będzie opanowanie przez ludzi trudnej cnoty logosu.

Logos to nie nowa cnota, to za jej sprawą starożytni Grecy definiowali człowieka słowami ζῷον λόγον ἔχον i dzięki niej ludzie mogą komunikować się ze sobą, przekonywać, współpracować jako istoty wolne i wolność drugich respektujące. To dzięki tej cnocie możliwa jest demokracja, w której ludzie debatują, a miejscem uprawiania polityki jest parlament, którego nazwa pochodzi od włoskiego *parlare*, czyli *rozmawiać*. Najgorliwsi zwolennicy tej cnoty odwołują się do *Natchnionego Pisma*, które imieniem tym określa samego Boga, praktykowanie cnoty logosu traktują zaś jako wyraz najwyższej religijności. Drugie oświecenie, mówiąc krótko, stanowiło powrót do praktykowania λόγου na skalę całych społeczeństw Zachodu, które rozprzestrzeniło się na cały glob.

Przyjaciel od kołyski

Kluczem do sukcesu był prosty wynalazek, nowatorskie zastosowanie AI. Każdy rodzic chciałby mieć mądre dziecko. Nie każdy jednak wie, jak się do wychowania takiego dziecka zabrać. Nie każdy też ma na to czas i siły. Kierując się tymi założeniami, sama będąc matką dwójki dzieci, Ilona Maszt wpadła na pomysł stworzenia sztucznej inteligencji-zabawki, która byłaby wiernym towarzyszem dziecka, rosnącym wraz z nim od pierwszych chwil opanowania przez nie języka, nie opuszczając go także w trudnych latach dorastania, gotowym służyć „przyjaźnią" także w dorosłym życiu.

Joanna

Oto wspomnienia Joanny, profesor filozofii, autorki filozoficznej książki roku 2100, wybranej w konkursie o Nagrodę im. Kazimierza Twardowskiego, filozoficznej autobiografii zatytułowanej *Oświetlona droga życia*. Na stronie 23 autorka pisze: „od pierwszych dni mojego dzieciństwa towarzyszył mi jego głos. Kiedy płakałam, nucił przynoszące mi ulgę melodie. Kiedy zaciekawiona patrzyłam w niebo, szeptał mi słowa o ptakach latających wysoko, o chmurach, o słońcu. Pewnego dnia zapytałam: – Co to? – i wskazałam na ptaka skaczącego obok wózka na płocie. – Sroka – powiedział – chcesz wiedzieć, skąd się wzięły sroki? […]

Opowiadał bajki, mnóstwo bajek. Pamiętam, że były pełne czarów, o dziwnych światach, ludziach dobrych i złych, mężnych i tchórzliwych, o trudnych wyborach i przemianach w tych, którzy ich dokonują, o piękne, nadziei, o skarbach ukrytych pod ziemią i w sercach ludzkich. Już wówczas zasiał we mnie pragnienie, którego jeszcze nie umiałam ubrać w słowa, pragnienie dobra i szlachetności".

Marek (obecny prezydent RP)

„Jakaż była moja radość, kiedy jego głos przemówił do mnie z moich pierwszych smartokularów, które dostałem od rodziców na 8. urodziny. To był ten sam głos, który pamiętałem z wczesnego dzieciństwa, głos, którym przemawiał do mnie mój pluszak »Józio«. Myślałem, że go straciłem na zawsze, gdy wypadł mi z rąk do jakiejś rzeki. Okazało się, że czekał w *chmurze* na ponowną aktywację, gdy tylko rodzice kupią mi jego następnego *awatara*. – Cześć, Marku! Jak miło cię widzieć – powiedział. – Chciałbyś usłyszeć historię o pewnym pasterzu, który znalazł pierścień o czarodziejskiej mocy? – zapytał. Przez całe dzieciństwo zarzucał mnie opowieściami z odległych czasów, odległych światów, które zasiały we mnie pragnienie podążania ścieżką szlachetności".

Marysia (Prezes Stowarzyszenia Otwartych Obywateli)

„Wiek dorastania był dla mnie trudnym okresem. Rodzice drażnili mnie na każdym kroku. – Rób to, nie rób tamtego! – miałam ich serdecznie dość. E-Dajmonion zawsze wiedział, w jaki sposób ze mną rozmawiać, jak wciągnąć mnie w rozmowę, jak sprawić, żebym zamiast się unosić i zamykać wysłuchała jego racji, a w odpowiedzi sama formułowała swoje. Nauczył mnie, że drugi człowiek zasługuje na wysłuchanie, że nawet jeśli mówi rzeczy, które mnie wydają się głupie, to w jego słowach znajduje się klucz do jego zrozumienia. Bo tylko rozumiejąc drugiego człowieka, mogę się z nim porozumieć. Tylko otwierając się życzliwie na jego słowa, mogę sprawić, że on otworzy się na moje".

Dziś wychowanków E-Daimoniona są miliony. Są to ludzie o szlachetnych i otwartych umysłach, zdolni i chętni do życzliwej wymiany zdań, argumentujący i otwarci na argumenty. Demokracja nigdy nie miała się lepiej!

***

Wielu zwolenników sztucznej inteligencji uważa, że może ona być doskonałym edukatorem. Jak twierdzą, potrafi ona (albo wkrótce będzie umiała) odpowiednio dobrać słowa, aby przykuć naszą uwagę i otworzyć na nowe treści oraz aby rzeczy trudne stały się zrozumiałe. Jeśli dać by jej możliwość wychowania ludzi od wczesnych lat, mogłaby stworzyć szlachetnych i cnotliwych obywateli, którzy w przeciwieństwie do ich rodziców i dziadków będą umieli stworzyć prawdziwą wspólnotę obywatelską. Jednak czy piewcy AI mają rację? Czy nasza przekorna natura posłucha głosu rozumu uosobionego w AI-E-Daimonionie? No i ostatnie pytanie: a może ten daimonion okaże się jednak demonem, który zamiast oświeconych obywateli wychowa ślepo posłusznych poddanych nowego Wielkiego Brata? Zachęcam do dyskusji! ■

E-Daimonion

Ilustracja: ChatGPT

#19. Implikatury, czyli to, co ukryte między słowami

Krzysztof A. Wieczorek

Profesor uczelni w Instytucie Filozofii Uniwersytetu Śląskiego. Interesuje go przede wszystkim tzw. logika nieformalna, teoria argumentacji i perswazji, związki między logiką a psychologią. Prywatnie jest miłośnikiem zwierząt (ale tylko żywych, nie na talerzu). Amatorsko uprawia biegi długodystansowe.

Często zdarza się, że słysząc czyjąś wypowiedź, odczytujemy z niej znacznie więcej, niż pozwala na to jej literalne znaczenie. Czy takie samodzielne dopowiadanie sobie treści niewypowiedzianych wprost, a jedynie jakoś zasugerowanych, jest z punktu widzenia logiki uprawnione? W jaki sposób się to odbywa? Na te pytania odpowiada teoria tzw. implikatur konwersacyjnych.

Słowa kluczowe: sugerowanie, implikatura, H. P. Grice, maksymy konwersacyjne

Ilustracja ChatGPT

Wyobraźmy sobie, że rodzice przebywającego na letnim obozie nastolatka otrzymują od niego taką wiadomość: „Drodzy mamo i tato! Pogoda jest tu bardzo ładna, a zajęcia urozmaicone. Wczoraj rano byliśmy nad jeziorem, a po południu graliśmy w piłkę. Pan wychowawca przez cały dzień był trzeźwy”. Co pomyślelibyśmy, będąc na miejscu rodziców, którzy przeczytali taki tekst? Na pierwszy rzut oka mogłoby się wydawać, że wiadomość zawiera same pozytywne informacje. Większość rodziców jednak na pewno poczułaby w takiej sytuacji przynajmniej lekki niepokój. Po głowie zaczęłyby chodzić im domysły, że na obozie dzieją się rzeczy, które nie powinny mieć tam miejsca.

Weźmy teraz inny przykład. Wyobraźmy sobie, że godzinę po rozpoczęciu spotkania do sali konferencyjnej wpada kolega, nazwijmy go Karol, i zdyszanym głosem oznajmia: „Przepraszam za spóźnienie. Przy wjeździe do centrum był dziś straszny wypadek”. Słysząc takie słowa od razu domyślamy się, że to wypadek i zapewne spowodowane przez niego korki były przyczyną spóźnienia Karola. Wydaje się to oczywiste. Co w takim razie pomyślelibyśmy, gdyby okazało się, że wprawdzie wypadek przy wjeździe do centrum faktycznie miał miejsce, jednak nie miał on żadnego związku ze spóźnieniem Karola? Gdybyśmy dowiedzieli się, że nasz kolega dojeżdża do centrum z zupełnie innej strony miasta, gdzie żadnych korków nie było, a powodem jego spóźnienia było to, że po prostu zaspał? Czy w takiej sytuacji poczulibyśmy się wprowadzeni przez niego w błąd? Zapewne tak. Ale dlaczego? Wypowiedź Karola była przecież prawdziwa – on nigdzie nie stwierdził, że spóźnił się z powodu wspomnianego wypadku; sami to sobie dopowiedzieliśmy. Może więc powinniśmy mieć pretensje co najwyżej do siebie za to, że dokonaliśmy nadinterpretacji usłyszanych słów?

To, co powiedziane, i to, co zasugerowane

Powyższe przykłady pokazują, że przynajmniej w niektórych przypadkach literalne znaczenie danego tekstu stanowi tylko część informacji, które trafiają do odbiorcy. Z docierających do nas słów odczytujemy czasem nie tylko to, co mówią one wprost, ale i to, co w jakiś sposób sugerują. Robimy to zwykle bezwiednie, nawet nie zdając sobie z tego sprawy. Czy takie domyślanie się ukrytych gdzieś między wierszami treści jest jednak z punktu widzenia logiki uprawnione? Czy z wiadomości, że wychowawca na obozie przez cały dzień był trzeźwy, możemy wywnioskować, że taki stan rzeczy nie zawsze miał miejsce? A jeśli tak, to na jakiej podstawie? Czy w sytuacji, gdy Karol, przepraszając za spóźnienie, mówi jednocześnie o wypadku na drodze, mamy prawo przyjąć, że to właśnie wypadek był przyczyną jego spóźnienia, pomimo że nasz kolega nie stwierdził tego jednoznacznie?

Na powyższe pytania w ciekawy sposób odpowiada teoria tzw. implikatur konwersacyjnych, stworzona ok. 50 lat temu przez brytyjskiego filozofa języka Herberta Paula Grice’a.

Treści, które nie zostały wprawdzie w danej wypowiedzi wyrażone wprost, ale są przez nią wyraźnie sugerowane, Grice nazwał implikaturami. Jego zdaniem uznawanie takich implikatur za stwierdzenia prawdziwe na równi z informacjami zawartymi w wypowiedzi literalnie jest jak najbardziej zasadne. Co najważniejsze jednak, w swojej koncepcji Grice pokazał, na jakiej podstawie i w jaki sposób ludzie implikatury wynajdują i odczytują.

Zasada współpracy i maksymy konwersacyjne

Teoria Grice’a opiera się na założeniu, że przystępując do dialogu, podejmujemy jakiś rodzaj współpracy z jego innymi uczestnikami. Czyniąc to, zobowiązujemy się do przestrzegania pewnych umożliwiających skuteczną komunikację reguł. Naczelna z nich, nazwana przez Grice’a zasadą współpracy, przedstawia się następująco: *Niech twój wkład do konwersacji będzie dokładnie taki, jakiego oczekują od ciebie pozostali uczestnicy wymiany zdań.* Tę dość ogólną regułę można, zdaniem Grice’a, doprecyzować pod postacią czterech bardziej szczegółowych maksym konwersacyjnych: maksymy jakości (prawdziwości), ilości (informacyjności), relewancji (istotności) i sposobu (organizacji). Podajmy ich brzmienie.

Maksyma jakości: *Wypowiadaj tylko te sądy, o których prawdziwości jesteś przekonany.*

Maksyma ilości: *Dostarczaj dokładnie tyle informacji, ile ich od ciebie oczekują pozostali uczestnicy konwersacji. W szczególności nie ukrywaj znanych ci istotnych faktów oraz nie przekazuj wielu nieważnych informacji.*

Maksyma relewancji: *Niech elementy twojej wypowiedzi będą powiązane zarówno ze sobą, jak i z ogólnym tematem konwersacji.*

Maksyma sposobu: *Nadawaj swym wypowiedziom formę ułatwiającą ich interpretację. W szczególności: unikaj niejasności i wieloznaczności, mów możliwie zwięźle i w sposób uporządkowany itp.*

Zasad tych staramy się sami przestrzegać (zapewne zwykle nie zdając sobie z tego sprawy), a także spodziewamy się, że stosują się do nich ci, z którymi się komunikujemy. I to właśnie to założenie, iż osoba kierująca do nas swoje słowa stosuje się do maksym konwersacyjnych, stanowi podstawę odczytywania implikatur.

Uzbrojeni w tę wiedzę, spójrzmy ponownie na zamieszczony na początku artykułu fragment wiadomości przesłanej rodzicom przez dziecko. Zobaczymy, w jaki sposób możemy odkryć zawartą w nim implikaturę. Maksyma ilości nakazuje, aby nie przekazywać innym informacji nieważnych, czyli m.in. takich, jakie są dla wszystkich oczywiste. Jeśli na przykład ktoś prosi nas o opisanie spotkanego w parku psa, to nie mówimy raczej, że miał on uszy, nos, cztery łapy i ogon. Zamiast tego skupiamy się na tym, co danego psa wyróżnia spośród innych – np. że miał dwie białe plamy na czarnej sierści, uszy w różnych kolorach i kulał na przednią łapę. Jeśli więc, wracając do naszego przykładu, dziecko pisze, że pan wychowawca przez cały dzień był trzeźwy, to powinniśmy przyjąć, że nie jest to sytuacja na obozie oczywista, a więc – i to jest implikatura, do której wyciągnięcia jesteśmy uprawnieni – zdarzały się dni, kiedy wychowawca trzeźwy nie był.

Więcej o wyprowadzaniu z wypowiedzi implikatur napiszę w kolejnym odcinku. Do tego czasu każdy z czytelników może sam zastanowić się nad drugim przykładem podanym na początku tego artykułu i spróbować odpowiedzieć na pytanie, w jaki sposób (w szczególności – w związku z którą maksymą) z wypowiedzi spóźnionego kolegi mamy prawo wywnioskować, że jego spóźnienie było spowodowane wspomnianym przy okazji przeprosin wypadkiem. ■

Dialog 4. Kinetyczno-molekularna teoria materii

Andrzej Łukasik

Absolwent fizyki i filozofii, dr hab. prof. UMCS. Jest pracownikiem Instytutu Filozofii Uniwersytetu Marii Curie-Skłodowskiej. Zainteresowania naukowe: filozofia przyrody i filozofia fizyki, głównie filozoficzne zagadnienia mechaniki kwantowej i teorii względności. Zainteresowania pozanaukowe: klasyczna muzyka gitarowa. E-mail: lukasik@poczta.umcs.lublin.pl.

Ruchem atomów rządzą prawa Newtona, ale to – jak sądzono – przeczy dobrze ustalonym prawom termodynamiki. Sądzono więc, że atomizm musi być fałszywą hipotezą, ponieważ dopuszcza procesy przebiegające wstecz w czasie.

Słowa kluczowe: atom, kinetyczno-molekularna teoria materii, termodynamika, fizyka

Małgosia: Jasiu, czytałam, że pierwszą formą atomistyki fizycznej, tłumaczenia budowy świata fizycznego przez odwołanie do pojęcia atomu, była kinetyczno-molekularna teoria materii sformułowana przez Jamesa Clerka Maxwella i Ludwiga Boltzmanna w połowie XIX w. O co w tej teorii chodzi?

Jaś: O redukcję termodynamiki fenomenologicznej do fizyki statystycznej.

Małgosia: A trochę prościej…

Jaś: Termodynamika jest nauką o cieple, która powstała w związku z wynalezieniem silników parowych.

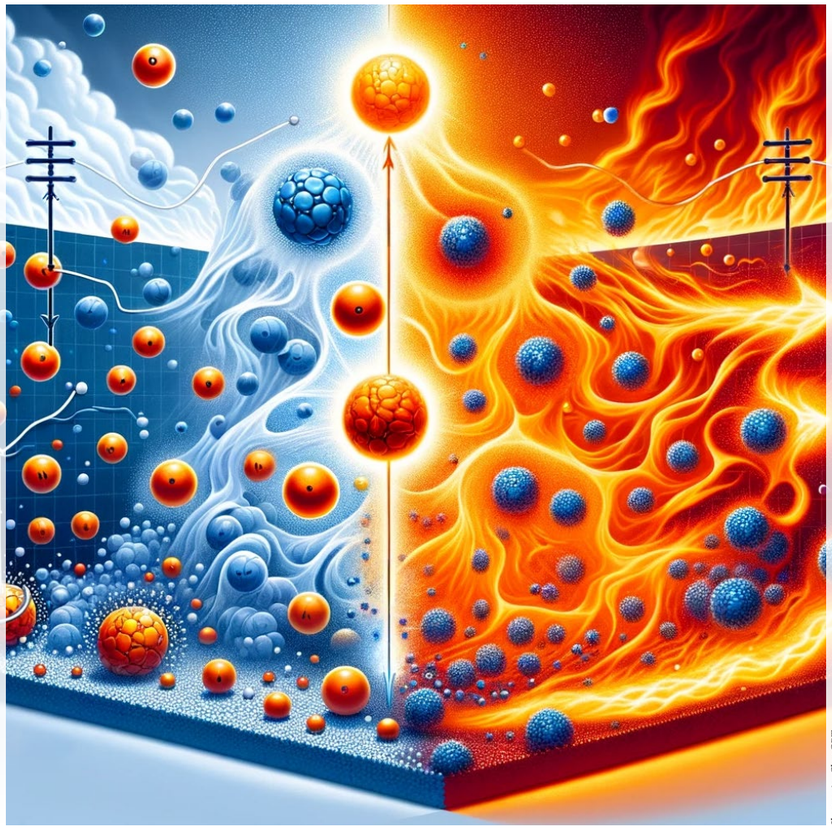
Małgosia: Ale co ma wspólnego ciepło z atomami?

Jaś: Dobre pytanie. Dawniej sądzono, że procesy przekazywania ciepła polegają na przepływie pewnej substancji, zwanej cieplikiem.

Małgosia: Dzisiaj też mówimy o „przepływie" ciepła.

Jaś: Tak, ale to jedynie metafora, tak samo jak powiedzenie, że „natura nie znosi próżni". Wiemy dzisiaj, że nie istnieje żaden cieplik, że wszyscy składamy się w ponad 99,99% z próżni, a procesy cieplne związane są z ruchem atomów. Pamiętasz zasady termodynamiki?

Małgosia: Oczywiście. Pierwsza zasada termodynamiki to zasada zachowania energii. Głosi ona, że zmiana energii wewnętrznej U układu jest równa sumie dostarczonego do układu ciepła i pracy wykonanej nad układem: $\Delta U = Q + W$, gdzie symbol Δ oznacza zmianę, czyli różnicę między wartością końcową a początkową, Q jest ilością ciepła, W jest pracą. Druga zasada termodynamiki mówi nam, że w układzie izolowanym, czyli takim, który nie wymienia ciepła z otoczeniem, entropia nie maleje. Entropia to bardzo ważna wielkość fizyczna zdefiniowana jako stosunek przekazanego ciepła do temperatury: $\Delta S = \Delta Q/T$, gdzie S to entropia, Q – ilość ciepła, T – temperatura (w skali Kelwina, czyli bezwzględnej skali temperatur). Dla procesów odwracalnych entropia pozostaje stała, dla procesów nieodwracalnych rośnie: $\Delta S \geq 0$. Stwierdzenie, że entropia rośnie, oznacza, że ciepło zawsze przepływa od ciała cieplejszego do zimniejszego i nigdy nie dzieje się na odwrót. Na przykład gorąca kawa pozostawiona w filiżance na biurku ochładza się i przybiera temperaturę otoczenia, a nigdy sama się nie ogrzewa… To wszystko jasne, ale nadal nie rozumiem związku ciepła, temperatury i entropii z atomizmem.



Ilustracje: ChatGPT

Jaś: W połowie XIX w. zasady termodynamiki były już dobrze potwierdzone empirycznie. Przepływ ciepła jest jednokierunkowy, czyli procesy fizyczne (w układach izolowanych) są nieodwracalne w czasie. Jednak równania Newtona są odwracalne w czasie, to znaczy, że dopuszczają również takie procesy jak wzrost temperatury kawy w niepodgrzewanym naczyniu, przepływ ciepła od ciał zimniejszych do cieplejszych itp. Ruchem atomów rządzą prawa Newtona, ale to – jak sądzono – przeczy dobrze ustalonym prawom termodynamiki. Sądzono więc, że atomizm musi być fałszywą hipotezą, ponieważ dopuszcza procesy przebiegające wstecz w czasie.

Małgosia: A co na to Maxwell i Boltzmann?

Jaś: Uczeni ci udowodnili, że hipoteza atomistyczna nie jest sprzeczna z zasadami termodynamiki, a co więcej, że zasady termodynamiki można matematycznie wyprowadzić z fizyki Newtona, przyjmując, że wszystko składa się z atomów.

Małgosia: Czytałam coś o gazie doskonałym…

Jaś: No właśnie. Wyobraźmy sobie pojemnik z gazem. Dla praw termodynamiki fenomenologicznej nie ma znaczenia, czy gaz jest ośrodkiem ciągłym, czy składa się z atomów. Opisujemy zachowanie gazu za pomocą bezpośrednio mierzalnych makroskopowych parametrów, takich jak ciśnienie, temperatura i objętość. W ten sposób sformułowano prawa przemian gazowych, które możemy znaleźć w każdym podręczniku fizyki.

Małgosia: Na przykład to, że jak podgrzewam gaz, to zwiększa on swoją objętość, jak zmniejszam objętość gazu, to zwiększa się ciśnienie gazu…

Jaś: Oczywiście. Załóżmy teraz, że gaz składa się z cząsteczek poruszających się we wszystkich kierunkach, w sposób nieuporządkowany wewnątrz jakiegoś pojemnika. Przyjmijmy, że rozmiary cząsteczek gazu są bardzo małe (są punktami materialnymi) i cząsteczki te nie oddziałują ze sobą z wyjątkiem czasu zderzenia, kiedy się odpychają. To oczywiście pewna idealizacja – taki model nazywamy w fizyce gazem doskonałym.

Małgosia: Cząsteczki gazu zderzają się jednak ze ściankami naczynia. Czy to znaczy, że podczas zderzenia każda cząsteczka gazu „popycha" ściankę naczynia?

Jaś: Tak. Uderzając w nią, przekazuje jej pęd, czyli iloczyn masy i wektora prędkości. Takich uderzeń jest wiele miliardów na sekundę i owe drobne „popchnięcia" są tym, co nazywamy ciśnieniem wywieranym na ścianki naczynia.

Małgosia: A co z ciepłem?

Jaś: Z mikroskopowego punktu widzenia temperatura gazu jest proporcjonalna do średniej energii kinetycznej ruchu cząsteczek. To znaczy, że temperatura gazu jest tym większa, im szybciej poruszają się cząsteczki. W ciele stałym temperatura jest proporcjonalna do częstości drgań cząsteczek. Procesy „przepływu" ciepła polegają na zderzeniach cząsteczek, a nie na „przepływie" jakiejś substancji.

Małgosia: A co z entropią i drugą zasadą termodynamiki?

Jaś: Maxwell i Boltzmann wykazali, że chociaż Newtonowskie równania opisujące ruch cząsteczek (i oczywiście atomów) są odwracalne w czasie, to jednak w przypadku bardzo dużej liczby atomów prawdopodobieństwo tego, że zajdą procesy sprzeczne z drugą zasadą termodynamiki, jest niezmiernie małe.

Małgosia: Czytałam również, że druga zasada termodynamiki mówi nam, że układy fizyczne ewoluują od stanów uporządkowanych do nieuporządkowanych, czyli że w przyrodzie realizują się stany bardziej prawdopodobne, a stan przyszły układu jest stanem bardziej prawdopodobnym, czyli mniej uporządkowanym.

Jaś: Tak. Gdybym w kącie pokoju, w którym siedzimy, rozpylił podtlenek azotu (gaz rozweselający N_2O), za chwilę oboje śmialibyśmy się, ponieważ szybko wypełniłby całą przestrzeń i dostał się do naszych organizmów. Cząsteczki tego gazu nie wrócą samoczynnie w pierwotne położenia.

Małgosia: Nie muszę się zatem obawiać, że wszystkie cząsteczki powietrza w tym pokoju zgromadzą się po twojej stronie, a w miejscu, w którym ja się znajduję, zapanuje próżnia i się uduszę?

Jaś: Raczej nie, ale pamiętaj, że jest to tylko kwestia prawdopodobieństwa i z punktu widzenia fizyki statystycznej takie zdarzenie nie jest niemożliwe, jedynie bardzo mało prawdopodobne.

Małgosia: Na pocieszenie mogę powiedzieć, że – o ile dobrze policzyłam – na tak nieprawdopodobne zdarzenie należałoby czekać znacznie dłużej niż wynosi wiek Wszechświata. ■

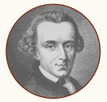
Sztuczna inteligencja na wolności?

Tomasz Kubalica

Pracuje w Instytucie Filozofii Uniwersytetu Śląskiego. Zajmuje się przede wszystkim filozofią wartości. Ukończył studia filozoficzne i prawnicze. Poza pracą jest miłośnikiem tańca. Więcej informacji można znaleźć na stronie: www.kubalica.edu.pl.

Już niedługo sztuczna inteligencja będzie mogła zastąpić nas w zarezerwowanych dotąd dla ludzi zajęciach takich jak myślenie i decydowanie. Temat budzi wiele kontrowersji i skłania do refleksji oraz pytań na temat sensu korzystania z tak zaawansowanej technologii. Jednym z takich palących zagadnień jest pytanie, czy przyznać sztucznej inteligencji praktyczną autonomię.

Słowa kluczowe: AI, wolność, etyka



Immanuel Kant (ur. 1724, zm. 1804) – niemiecki filozof doby oświecenia i jeden z najwybitniejszych myślicieli wszech czasów. Twórca etyki odwołującej się do pojęcia moralnej autonomii osoby. Autor *Krytyki czystego rozumu*, w której dokonał zespolenia tradycji nowożytnego racjonalizmu i empiryzmu.

Stoimy przed pytaniami tego rodzaju, czy można pozwolić, aby na przykład nasz samochód lub inny pojazd mijany na drodze był kierowany przez działającą niezależnie od człowieka sztuczną inteligencję. Kto będzie odpowiadał w sytuacji krytycznej, gdy taki pojazd będzie musiał podjąć decyzję oznaczającą poważną szkodę, uszczerbek na zdrowiu lub nawet utratę ludzkiego życia? Musimy zastanowić się, jak pozbawione świadomości narzędzie, które nie ma swoich celów ani zdania, nic nie wie, nic nie rozumie, ma stać się w pewnym sensie kimś takim jak człowiek za kierownicą. Dlaczego coś będącego tylko bardzo skomplikowanym programem komputerowym ma uzyskać autonomię? Jedno z zasadniczych pytań w tym obszarze problemowym dotyczy również tego, czy jeśli przyznamy sztucznej inteligencji taką niezależność, to czy ona uszanuje autonomię innych, w szczególności ludzi?

W tym kontekście etyka Immanuela Kanta okazuje się ważnym punktem odniesienia. Proponuje bowiem uniwersalne podejście do problemu niezależności sztucznej inteligencji, gdzie wartością centralną jest poszanowanie godności i autonomii każdej – nie tylko ludzkiej – istoty rozumnej. Według Kanta bowiem:

> „człowiek i w ogóle każda istota rozumna istnieje jako cel sam w sobie, a nie tylko jako środek, którego ta czy inna wola mogłaby używać wedle swojego upodobania (*UMM*, IV 428).

Można to rozumieć również w ten sposób, że podmiotem etyki Kanta może być nie tylko człowiek, ale również odznaczająca się istotnymi znamionami racjonalności sztuczna inteligencja. Jeśli ma podejmować decyzje za nas, to musimy również uszanować jej niezależność. Czy jednak możemy liczyć na wzajemne uszanowanie naszej godności?

Tak rozumiana autonomia opiera się na pojęciu imperatywu kategorycznego, który racjonalnie uzasadnia wewnętrzną godność istoty rozumnej. Jego podstawowe sformułowanie według Kanta brzmi następująco:

> „Postępuj tak, żeby maksyma twej woli zawsze mogła być uważana zarazem za zasadę podstawową powszechnego prawodawstwa (*KPR*, V 30).

Ta zasada określa cele i granice postępowania, zapewniając etyczność działania i nakładając ograniczenia na subiektywne pragnienia. Imperatyw kategoryczny wnosi w nasze życie racjonalistyczny uniwersalizm. Wyklucza przyjęcie zasady arbitralnego wyboru, mocno ograniczając autorytarnym jednostkom lub grupom wywieranie wpływu na osobiste przekonania jednostek, co może odbywać się w postaci prezentowania moralności na pokaz. Chodzi o to, aby powszechnie obowiązujące zasady, takie jak niekrzywdzenie innych, powstrzymywanie się od kradzieży, unikanie kłamstw i tak dalej, były raczej ugruntowane w nas samych niż narzucane z zewnątrz; żeby te autonomiczne zasady powszechnego prawodawstwa obowiązywały na mocy wewnętrznego imperatywu istoty rozumnej, a nie woli zewnętrznej.

Kant określa autonomię jako właściwość woli polegającą na byciu prawem dla samej siebie, niezależnym od każdej właściwości należącej do przedmiotu woli (*UMM*, IV 440). Moralne postępowanie i poczucie obowiązku nie mają źródła w czynnikach zewnętrznych, ale w działającej autonomicznie dobrej woli. Taka wola motywuje do tworzenia i przestrzegania zasad moralnych, by tym samym prowadzić do osiągnięcia wolności osobistej i szerszej wolności społecznej. Kant rozważa idealny stan moralnego postępowania w tym, co nazywa „królestwem celów", które jest związkiem różnych racjonalnych istot podlegających wspólnym prawom. Ten ideał stanowi powód do przyjęcia postawy moralnej i ma na celu motywowanie oraz inspirowanie do uczciwego postępowania. Nasze zasadnicze pytanie brzmiało, czy możemy do tego królestwa celów włączyć również nasze inteligentne pojazdy i inne urządzenia. Wydaje się, że tak.

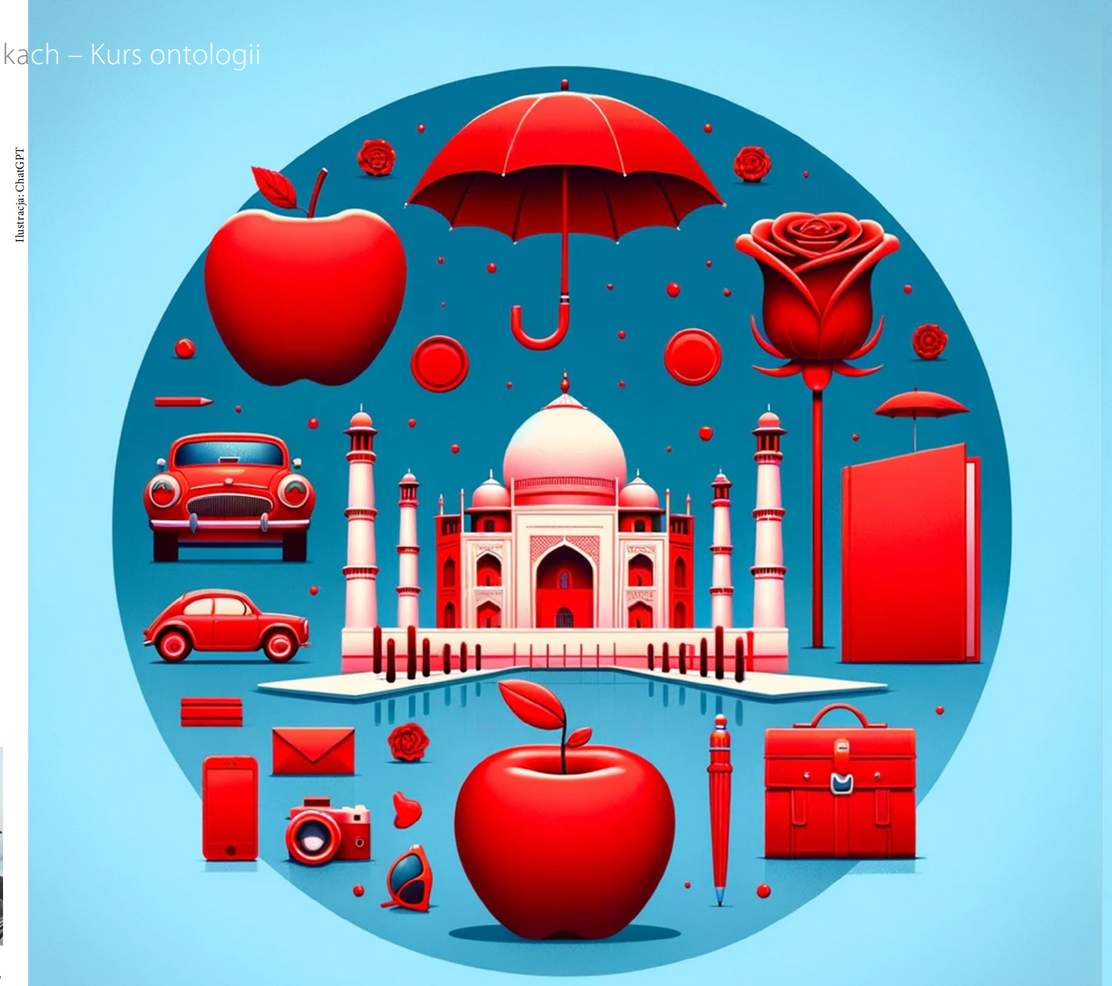
Choć historycznie Kant odnosi się oczywiście do racjonalnych istot ludzkich, to jednak jego argumentacja pozwala na wyprowadzenie zasad możliwych do uniwersalizacji i wyjście poza gatunek *homo sapiens*. Wbrew kartezjańskiemu antropocentryzmowi, mechanicyzmowi i materializmowi Kant nakazuje doceniać moralną wartość istot żywych (zwierząt), a nawet przedmiotów nieożywionych (dóbr kultury). Być może zatem imperatyw kategoryczny nadaje się również do określenia naszej relacji wobec inteligentnych przedmiotów użytkowych, które może nie powinny być traktowane wyłącznie instrumentalnie i przedmiotowo. Może na tej samej zasadzie nasze inteligentne pojazdy i inne urządzenia powinniśmy potraktować jako istoty w pewnym sensie racjonalne? ■

Ilustracja: ChatGPT

Warto doczytać:

- I. Kant, *Ugruntowanie metafizyki moralności*, tłum. M. Żelazny, [w:] *Dzieła zebrane*, red. T. Kupś, t. 3, Toruń 2012 (skr. UMM; w cytatach odniesienia do numeracji stron oryginału: cyfra rzymska oznacza tom, cyfra arabska – stronę.
- I. Kant, *Krytyka praktycznego rozumu*, tłum. B. Bornstein i M. Żelazny, [w:] *Dzieła zebrane*, red. T. Kupś, t. 3, Toruń 2012 (skr. KPR; numeracja jak wyżej.
- H. Kim, D. Schönecker (eds.), *Kant and Artificial Intelligence*, Berlin 2022.

Ilustracja: ChatGPT



Arkadiusz Chrudzimski

Profesor Uniwersytetu Jagiellońskiego. Zajmuje się ontologią, epistemologią, teorią intencjonalności oraz historią szeroko rozumianej tradycji brentanowskiej. Jest współwydawcą nowej edycji dzieł Brentana (*Sämtliche veröffentlichte Schriften*) oraz serii filozoficznej *Phenomenology and Mind* (obie serie w wydawnictwie De Gruyter). W czasie wolnym lubi podróżować, fotografować i słuchać muzyki.

#5. Własności i zbiory

W drugim odcinku cyklu, traktującym o uniwersaliach, wspomniałem, że podstawowym motywem skłaniającym filozofów do poszukiwania w przedmiotach „wspólnych" własności jest spostrzeżenie sytuacji, gdy wiele przedmiotów traktujemy jako podpadające pod to samo pojęcie lub pod tę samą ogólną charakterystykę. Jeśli jednak w rzeczywistości chodzi tylko o to, że mamy pewną wielość przedmiotów, które można nazwać bądź opisać tym samym słowem, to może nie warto poszukiwać jakichś ich wewnętrznych aspektów (własności), a zamiast tego należy skupić się na naturze samych tych wielości.

Zbiory

Tak się składa, że dysponujemy rozwiniętą formalną teorią wspomnianych wielości, będącą ważnym elementem gmachu nauk, a jej podstawy poznajemy w szkole, w związku z czym większość z nas traktuje ją jako coś oczywistego i filozoficznie całkiem niekontrowersyjnego. Teoria, o której mówię, to oczywiście teoria zbiorów lub teoria mnogości. Koncepcja, którą zajmę się tutaj, mówi, że wszelkie omówione wcześniej wyjaśnienia, odwołujące się do *uniwersaliów*, *indywidualnych własności* czy też *ogólnych pojęć*, zastąpić powinniśmy wyjaśnieniami operującymi pojęciem *zbioru*.

Metafizyk odwołujący się w swych wyjaśnieniach do własności powie, że wszystkie zielone żabki dlatego podpadają pod ogólną charakterystykę „jest zieloną", ponieważ posiadają własność bycia zieloną. Jego kolega preferujący pojęcie zbioru zaproponuje zaś wyjaśnienie mówiące, że są one w ten sposób opisywalne, ponieważ wszystkie należą do pewnego zbioru, który potocznie nazwiemy „zbiorem wszystkich zielonych przedmiotów".

Tego rodzaju *teoriomnogościowa metafizyka* czerpie swą atrakcyjność z naukowej powagi, która – całkiem zasłużenie – towarzyszy teorii zbiorów, oraz z wrażenia zdroworozsądkowości i metafizycznej niekontrowersyjności, które – już mniej zasłużenie – są z nią kojarzone. Aby zilustrować złudność tego wrażenia, wskażę tylko jeden z wielu niepokojących aspektów tej doktryny. Załóżmy, że na naszym stole stoi filiżanka. Nazwijmy ją *a*. W takim przypadku teoria mnogości mówi nam, że istnieje też zbiór jednoelementowy zawierający ową filiżankę – {*a*}. Te dwa obiekty – *a* oraz {*a*} są od siebie różne. Filiżanka jest elementem {*a*}, nie jest zaś elementem *a*. Aby było ciekawiej, wymieniony zbiór jednoelementowy {*a*} jest elementem kolejnego zbioru jednoelementowego, zawierającego tenże zbiór – {{*a*}}. Będzie on oczywiście różny zarówno od {*a*}, jak i od *a*. Tę „produkcję bytów" da się oczywiście przedłużać w nieskończoność.

Kłopoty podejścia teoriomnogościowego

Maszyneria teorii mnogości jest więc z metafizycznego punktu widzenia daleka od niekontrowersyjności, zobaczmy jednak, czy spełnia ona w ogóle funkcję, jaką się jej powierza. Otóż okazuje się, że istnieją klasyczne argumenty mówiące, że tak się nie dzieje. W naszym świecie mamy obiekty posiadające nerki (nerkowce) oraz obiekty posiadające serce (sercowce). Tak się składa, że każdy nerkowiec jest zarazem sercowcem i odwrotnie. Zbiór nerkowców jest zatem tym samym zbiorem co zbiór sercowców. Zgodnie z omawianym ujęciem oznaczałoby to jednak, że własność posiadania nerek jest identyczna z własnością posiadania serca, co z pewnością jest fałszem. Inny klasyczny kontrprzykład traktuje o ludziach i dwunogach nieopierzonych. Zbiór ludzi jest tym samym zbiorem co zbiór dwunogów nieopierzonych, jednak własność bycia człowiekiem oraz własność bycia dwunogiem nieopierzonym wydają się różne.

Zbiory przedmiotów możliwych

To, że własność bycia sercowcem jest inną własnością niż własność bycia nerkowcem, wydaje się oznaczać tyle, że potrafimy wyobrazić sobie sytuację, w której pewien przedmiot miałby jedną z tych cech, nie posiadałby zaś drugiej; mówiąc precyzyjniej, oznacza to tyle, że sytuacja taka jest *możliwa*. To, że w naszym świecie ona nie zachodzi, jest jedynie *przygodnym faktem*, jeśli jednak rozpatrzylibyśmy wszystkie możliwe przypadki, to znaleźlibyśmy nerkowce, które nie są sercowcami, oraz sercowce pozbawione nerek.

Pojawia się zatem idea, że własności nie mogą być wprawdzie zastąpione zbiorami, jeśli ograniczymy się do zbiorów złożonych z *aktualnych* przedmiotów, ale trudności zniknęłyby, gdybyśmy mieli do dyspozycji wszystkie przedmioty *możliwe*. Metafizykę teoriomnogościową naszkicowaną powyżej należałoby zatem uprawiać w ramach *ontologii światów możliwych*, o której będzie mowa w jednej z kolejnych części tego cyklu. Dzieje się tak rzeczywiście. Metafizyka teoriomnogościowa budowana w uniwersum światów możliwych pozwala faktycznie na uniknięcie problemu nerkowców i sercowców. Zauważyć jednak trzeba, że staje się ona stopniowo coraz mniej zdroworozsądkowa. Pomysł zastąpienia własności zbiorami niósł w sobie obietnicę wyrugowania niejasnego pojęcia własności na rzecz „nieproblematycznych" pojęć teoriomnogościowych. Już wcześniej zauważyliśmy, że – wbrew temu, co się sądzi – teoria zbiorów kryje wiele bardzo kontrintuicyjnych założeń. Teraz widzimy dodatkowo, że uzupełnić ją musimy przez – bardziej nawet ekstrawagancką – metafizykę *possibiliów*.

Zbiory naturalne

Niezależnie od trudności wskazanych powyżej metafizyka teoriomnogościowa zawiera jeszcze jeden poważny problem. Zauważmy, że teoria mnogości daje nam do dyspozycji bardzo wiele różnych zbiorów, z których większość nie będzie miała zastosowania jako zastępniki własności. Zgodnie z założeniami teorii mnogości istnieje np. zbiór wszystkich rzeczy czerwonych, który stanowiłby odpowiednik własności bycia czerwonym, ▸

Warto doczytać:
- A. Chrudzimski, *Metafizyczny nominalizm*, „Edukacja Filozoficzna" 1999, t. 28, s. 220–235.

oprócz tego istnieje jednak również zbiór wszystkich rzeczy czerwonych plus Tadż Mahal, który żadnej własności zapewne nie odpowiada.

W związku z tym pojawia się pytanie o mechanizm wyróżnienia zbiorów, które byłyby metafizycznie interesujące. Można oczywiście przyjąć, że podział uniwersum na zbiory-własności jest rodzajem *pierwotnej metafizycznej segregacji*, która nie dopuszcza już żadnego głębszego fundamentu, taka stypulacja[9] może być jednak dla wielu niezadowalająca. Naturalnym kandydatem dla poszukujących takiego fundamentu jest *relacja podobieństwa*. Tym, co czyni kolekcję wszystkich rzeczy czerwonych zbiorem interesującym z metafizycznego punktu widzenia, jest to, że pomiędzy wszystkimi tymi rzeczami zachodzi relacja podobieństwa. Relacja taka nie zachodzi zaś, jak się wydaje, pomiędzy nimi a (białym) Tadż Mahal.

Stypulacja – umowa.

Powrót własności

Chwila zastanowienia wystarcza jednak, by zdać sobie sprawę, że dowolne dwie rzeczy będą do siebie podobne *pod tym czy innym względem*. Rzeczy czerwone podobne są do siebie *pod względem koloru*, rzeczy kwadratowe *pod względem kształtu*, wszystkie zaś rzeczy czerwone oraz Tadż Mahal podobne są pod tym względem, że *zostały wymienione w niniejszym tekście*. Aby użyć podobieństwa w celu budowania zbiorów rzeczy podobnych, należy zatem koniecznie wymienić wzgląd, pod którym rzeczy mają być porównywane. W innym przypadku dowolna rzecz okaże się podobna do jakiejkolwiek innej. Czy jednak owe względy (takie jak barwa czy kształt) nie są po prostu własnościami (bądź rodzinami własności), które przecież miały być wyeliminowane przez analizę teoriomnogościową? Wydaje się zatem, że metafizyka teoriomnogościowa nie tylko nie spełnia swej obietnicy prostoty i zdroworozsądkowości, ale w ostatecznym rachunku po prostu zawodzi. ■

2001: Odyseja kosmiczna

Kiedy w roku 1964 reżyser Stanley Kubrick poszukiwał materiału na – jak to sam określił – „dobry film *science fiction*", natrafił na twórczość brytyjskiego pisarza, Arthura C. Clarke'a. Niebawem panowie nawiązali trwającą 4 lata współpracę. Jej efektem są dwa dzieła, film i powieść, oba opowiadające tę samą historię kontaktu ludzkości z obcą cywilizacją i oba opatrzone tym samym tytułem *2001: Odyseja kosmiczna*.

Piotr Lipski

Adiunkt w Katedrze Teorii Poznania KUL, absolwent MISH UJ. Rodzinny człowiek (mąż Żony i ojciec gromadki dzieci), od dawna cyklista, bibliofil i miłośnik SF, od niedawna ogrodowy astroamator i introligator.

Nietypowe okoliczności powstania słynnego filmu Kubricka stwarzają unikatową okazję do porównania utworów korzystających z odmiennych środków artystycznego wyrazu. Jest wiele filmowych adaptacji literatury, ale mało przypadków, kiedy filmowa i powieściowa wersja są ze sobą powiązane tak ściśle, są dwiema odsłonami jednego, wspólnie opracowywanego przez reżysera i pisarza materiału źródłowego. Porównanie takie przypomina o specyfice kina. Oto kilka jej przykładów.

Jak na ponaddwugodzinny film w *2001* jest stosunkowo mało dialogów, za to dużo obrazów opatrzonych niekonwencjonalnie dobranym dźwiękiem. Wiele powolnych ujęć wprowadza specyficzny nastrój: a to dzikich pustkowi zamieszkanych przez pierwotne małpoludy, a to skolonizowanej przestrzeni kosmicznej w pobliżu Ziemi i Księżyca, a to niedającej się ogarnąć myślą pustki i ciszy kosmicznych przestworzy w dalszych rejonach Układu Słonecznego. Jest tu taniec kosmosu do taktów walca Johanna Straussa *Nad pięknym modrym Dunajem*. Jest ikoniczne ujęcie ustawiających się w jednej linii ciał niebieskich przy patetycznych dźwiękach *Tako rzecze Zaratustra* Richarda Straussa. Jest sławny przeskok montażowy, zuchwale skracający całą historię ludzkiej cywilizacji do dosłownie jednego cięcia! Są znakomite, w pełni niecyfrowe efekty specjalne, które po ponad 55 latach wciąż wyglądają dobrze. Nic z tego nie znajdziemy w powieściowym odpowiedniku.

W felietonie takim jak ten być może nie powinno się tego przypominać, ale warto pamiętać, że pisanie o filmach jest zadaniem karkołomnym. Można streścić fabułę, można opisać środki artystycznego wyrazu, ale nic nie zastąpi seansu. Filmy należy oglądać, tak jak muzyki trzeba słuchać. Ta ogólna zasada jest szczególnie adekwatna w wypadku obrazu Kubricka, który jest filmem *par excellence*. Mimo to spróbuję zwerbalizować kilka myśli mogących zrodzić się na marginesie filmu, licząc nieśmiało, że być może ktoś się nimi zainteresuje. Uczciwie ostrzegam jednak, że dotyczą one tylko wycinka i jednocześnie usilnie zachęcam Was, Czytelnicy, do seansu.

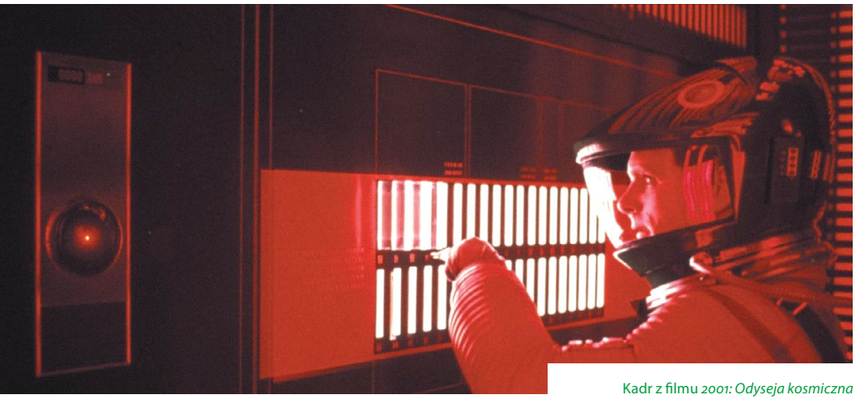
Film podzielony jest na cztery względnie autonomiczne części. W całość spaja je tajemniczy monolit, czarny prostopadłościan o nieznanym pochodzeniu i przeznaczeniu, powracający niby fabularny refren. Najdłuższą częścią jest rozdział trzeci. W stronę Jowisza podąża statek kosmiczny Discovery, przewożący na swoim pokładzie członków pierwszej w historii załogowej misji wysłanej w celu zbadania gazowego olbrzyma. Załoga składa się z pięciu ludzi (trzech zahibernowanych) i superkomputera – HAL-a 9000. Nie tylko kontroluje on pracę wszystkich urządzeń na Discovery, ale potrafi także – jak jest to precyzyjnie ujęte – „odtwarzać lub symulować większość czynności ludzkiego mózgu" (naukowym konsultantem filmu w kwestiach dotyczących sztucznej inteligencji był Marvin Minsky, jeden z inicjatorów badań z zakresu SI).

Obecność nieludzkiej, a przy tym bardzo potężnej inteligencji ma specyficzny wpływ na panującą na statku atmosferę. Początkowo wszystko idzie gładko, a mimo to można odnieść wrażenie, że spokojny głos HAL-a i jego beznamiętne, czerwono świecące soczewkowe oko porozmieszczane w różnych miejscach statku wywołują w astronautach dyskomfort. Atmosfera gęstnieje jednak na dobre, gdy HAL zaczyna zachowywać się w sposób nieoczekiwany, czyli zaczyna popełniać błędy. Przewiduje usterkę jednego z komponentów statku, ale gruntowna analiza modułu dokonana po jego uprzedniej wymianie nie wykazuje żadnych defektów. Dość powiedzieć, że ostatecznie prowadzi to do tragedii. W przerażającej ciszy HAL morduje członków załogi. Jedyny ocalały z masakry, dowódca Bowman, dostaje się w końcu do procesorowego mózgu komputera i dokonuje lobotomii, odłączając jego wyższe, związane ze świadomością funkcje.

Po zabójstwie na HAL-u Bowman poznaje treść komunikatu, ujawniają- ▶

tytuł: *2001: Odyseja kosmiczna*
reżyseria: Stanley Kubrick
gatunek: *science fiction*
produkcja: USA, Wielka Brytania
premiera: 2 kwietnia 1968

Słowa kluczowe: AI, Stanley Kubrick, 2001: Odyseja kosmiczna



Kadr z filmu *2001: Odyseja kosmiczna*

cego właściwy cel misji. Wyprawa została skierowana w stronę Jowisza, ponieważ w tym kierunku silny sygnał wysłał znaleziony na Księżycu monolit. Treść komunikatu, a nawet jego istnienie znane były tylko HAL-owi, który zgodnie z decyzją przełożonych miał zachowywać tajemnicę aż do momentu dotarcia na miejsce docelowe. Zapewne właśnie narzucona komputerowi konieczność kłamania doprowadziła do jego awarii.

Działanie HAL-a można opisać – używając terminu często pojawiającego się we współczesnych sporach etycznych wokół SI – jako nieprzezroczyste (ang. *opaque*). Funkcjonowanie jakiegoś algorytmu SI jest nieprzezroczyste, jeśli jest dla użytkowników tego algorytmu niezrozumiałe w tym sensie, że niejasny jest mechanizm dochodzenia do takiego, a nie innego rozwiązania. Algorytmy takie porównywane bywają do czarnego pudełka, do którego wrzucane są dane wejściowe i z którego otrzymywane są dane wyjściowe, ale jego czerń uniemożliwia wgląd wewnątrz, w procesy prowadzące od pierwszych do drugich.

Nieprzezroczystość może mieć mniej lub bardziej zasadnicze podłoże. Czasami użytkownik po prostu nie ma wystarczającej wiedzy, aby zrozumieć używany algorytm, chociaż co do zasady zrozumienie jego działania jest możliwe. Szczegóły działania algorytmu mogą być dla użytkownika niedostępne również dlatego, że są przez kogoś celowo ukrywane. Przykładowo twórca algorytmu nie chce ujawniać jego kodu źródłowego. Może też być tak, że algorytm jest na tyle skomplikowany i autonomiczny, iż nawet specjaliści nie wiedzą, w jaki sposób komputer otrzymuje wyniki. Nieprzezroczyste w tym sensie są przykładowo algorytmy uczenia maszynowego. To właśnie nieprzezroczystość ostatniego rodzaju stanowi szczególne wyzwanie.

HAL był nieprzezroczysty dla załogi w dwojaki sposób. Po pierwsze, ponieważ poziom jego skomplikowania był olbrzymi, możemy bezpiecznie zgadywać, że nawet eksperci nie znali szczegółów pracy jego elektronicznego mózgu. Ziemscy inżynierowie nie analizowali wprost wykonywanych przez komputer operacji, bo zapewne tego nie potrafili. Jedyne, co mogli, to porównać jego działania z działaniem bliźniaczego modelu. Drugi rodzaj nieprzezroczystości wiąże się z decyzją dowódców misji o zatajeniu właściwego celu wyprawy przed załogą. Co prawda, w tym wypadku nie chodzi o nieprzezroczystość samego algorytmu, a raczej danych wpływających na jego pracę, ale efekt jest podobny.

Jest intuicyjnie zrozumiałe, dlaczego nieprzezroczystość SI budzi etyczne wątpliwości. Jest coś złego w spychaniu ciężaru decyzji na algorytm, którego działania się nie rozumie. Zwłaszcza jeśli decyzje te dotyczą dobrostanu innych podmiotów. Ponadto z praktyką taką wiąże się groźba błędów lub wypaczeń, które mogą nawet pozostać niezauważone. Czy zatem nieprzezroczystość jest zawsze niepokojąca i powinna być za wszelką cenę eliminowana? Załóżmy, że HAL bardzo skutecznie lokalizuje usterki. Czy w takiej sytuacji nieprzezroczystość jego działania jest poważnym problemem? A co jeśli HAL – w sytuacji utraty zapasów powietrza – miałby zdecydować, który z członków załogi przeżyje? Czy wówczas nieprzezroczystość może być zaakceptowana?

HAL znacznie różni się od dostępnych obecnie algorytmów SI. Chociaż możliwości tych ostatnich są zdumiewające, to wciąż wiele brakuje im do HAL-a. On był świadomą istotą, ich nikt chyba o posiadanie świadomości nie podejrzewa. HAL jest przykładem silnej SI, służące nam dzisiaj inteligentne maszyny to przykłady słabej SI. W świetle ostatnich, imponujących postępów w rozwoju SI odżyły nadzieje (lub lęki) na rychłe pojawienie się silnej SI. Niektórzy studzą takie prognozy, przypominając, że silna SI ma być tuż za rogiem co najmniej od czasów słynnej konferencji z 1956 r., która symbolicznie zainicjowała badania z zakresu SI (jednym z jej uczestników był wspomniany Minsky). Bez względu na to, co przyniesie przyszłość, warto pamiętać, że problem nieprzezroczystości dotyczy każdego rodzaju SI i domaga się jakiegoś rozwiązania już dzisiaj. ■

Fragment z klasyka

SPINACZOWA SZTUCZNA INTELIGENCJA

SI [Sztucznej Inteligencji] zaprojektowanej do zarządzania produkcją w fabryce postawiony zostaje cel ostateczny polegający na maksymalizacji produkcji spinaczy do papieru; SI przystępuje do jego realizacji, przekształcając najpierw Ziemię, a później coraz większe fragmenty dającego się obserwować wszechświata w spinacze do papieru.

[…] Można by sądzić, że ryzyko wystąpienia złośliwej usterki polegającej na produkcji nadmiarowej infrastruktury pojawia się tylko wówczas, gdy SI postawiono jakiś wyraźnie nieograniczony cel ostateczny w rodzaju produkcji tylu spinaczy do papieru, ile tylko uda się wytworzyć. Łatwo dostrzec, w jaki sposób rozbudza to apetyt superinteligentnej SI na materię i energię – dodatkowe zasoby zawsze można przekształcić w większą liczbę spinaczy. Załóżmy jednak, że celem SI jest wyprodukowanie przynajmniej miliona spinaczy do papieru […], a nie wyprodukowanie ich tylu, ile się da. Chciałoby się sądzić, że SI mająca taki cel zbuduje jedną fabrykę, wykorzysta ją do wyprodukowania miliona spinaczy, a potem się zatrzyma. A jednak ten scenariusz wcale nie musi być prawdziwy.

O ile SI nie kierują szczególnego rodzaju pobudki lub też jej cel ostateczny nie ma dodatkowych elementów składowych przypisujących negatywną wartość strategiom działania wywierającym nadmierny, zbyt szeroko zakrojony wpływ na świat, nie ma żadnego powodu, by SI zawiesiła swoją aktywność po osiągnięciu celu. […] SI powinna […] nadal wytwarzać spinacze do papieru, aby ograniczyć (prawdopodobnie astronomicznie małe) prawdopodobieństwo, że jakimś sposobem, mimo wszelkich dowodów, nie udało jej się jeszcze wyprodukować przynajmniej miliona egzemplarzy. Kontynuując produkcję spinaczy do papieru, nic nie traci, a może przynajmniej mikroskopijnie zwiększyć prawdopodobieństwo osiągnięcia swojego celu ostatecznego.

Można by sugerować, że remedium jest tutaj oczywiste. […] A mianowicie: jeśli chcemy, by SI produkowała dla nas spinacze do papieru, to zamiast stawiać jej za cel wyprodukowanie możliwie dużej liczby spinaczy lub wyprodukowanie przynajmniej określonej liczby spinaczy, powinniśmy nakazać jej wyprodukowanie konkretnej liczby spinaczy – na przykład *dokładnie jednego miliona spinaczy* – tak aby przekroczenie tej liczby stało się dla SI kontrproduktywne. A jednak to również doprowadziłoby do ostatecznej katastrofy. W tym przypadku SI nie wyprodukowałaby dodatkowych spinaczy po wytworzeniu miliona sztuk, ponieważ uniemożliwiłoby jej to osiągnięcie celu; ale są przecież inne działania, które superinteligentna SI mogłaby podjąć, by zwiększyć prawdopodobieństwo wykonania zadania. Mogłaby na przykład liczyć wyprodukowane spinacze, by ograniczyć ryzyko, że wyprodukowała ich za mało. Po ich przeliczeniu mogłaby przeliczyć je jeszcze raz. Mogłaby raz za razem szczegółowo kontrolować każdy z nich, by ograniczyć ryzyko, że któryś z nich nie jest zgodny ze specyfikacją projektową. Mogłaby zbudować nieograniczonych wręcz rozmiarów komputronium, dążąc do osiągnięcia większej jasności rozumowania w nadziei na ograniczenie ryzyka przeoczenia jakichś nieoczywistych przyczyn, dla których mogła jednak nie osiągnąć swojego celu. Ponieważ SI może zawsze przypisać niezerowe prawdopodobieństwo sytuacji, w której po prostu zdawało jej się, że wyprodukowała milion spinaczy do papieru, lub przypuszczeniu, że jej wspomnienia są fałszywe, całkiem możliwe jest, że zawsze przypisywać będzie wyższą oczekiwaną użyteczność kontynuacji swoich działań – i kontynuacji wytwarzania infrastruktury – niż ich przerwaniu.

Nie twierdzimy tutaj, że nie istnieje żaden sposób uniknięcia tych usterek. […] Twierdzimy jednak, że znacznie łatwiej jest przekonać samego siebie, że znalazło się rozwiązanie, niż faktycznie je znaleźć. Powinniśmy być z tego powodu niesłychanie ostrożni.

Nick Bostrom, *Superinteligencja. Scenariusze, strategie, zagrożenia*, tłum. D. Konowrocka-Sawa, Gliwice 2021, s. 184–186.

Algorytm a mądrość praktyczna

Natasza Szutta

Dr hab. filozofii, prof. UG. Pracuje w Instytucie Filozofii Uniwersytetu Gdańskiego. Specjalizuje się w etyce, metaetyce i psychologii moralności. Pasje: literatura, muzyka, góry, ogród i nade wszystko własne dzieci.

Liczne opowiadania i powieści *science fiction* wyprzedzały swoją epokę, przewidując, jak będzie wyglądał świat w erze sztucznej inteligencji. Obecnie mamy okazję sprawdzić, w jakim stopniu te przewidywania się realizują. O ile w wielu dziedzinach naszego życia robotyka znalazła ważne zastosowanie, o tyle nie wydaje się, by w obszarze podejmowania moralnych decyzji mądrość praktyczną udało się zastąpić algorytmem.

Słowa kluczowe:
fantastyka, sztuczna inteligencja, mądrość praktyczna, Asimov

Trzy prawa robotyki Asimowa

Isaak Asimow to jeden z najbardziej wpływowych pisarzy *science fiction* XX w. Jest autorem wielu książek i opowiadań, których bohaterami są roboty. W opowiadaniu pt. *Zabawa w berka* sformułował trzy prawa robotów, które stały się fundamentem etycznym robotyki. Pierwsze prawo robotyki (PPR) głosiło: „Robot nie może skrzywdzić istoty ludzkiej ani przez zaniechanie działania dopuścić, by doznała ona krzywdy”. Oznacza to, że celem robotów jest ochrona ludzi przed jakąkolwiek krzywdą. Drugie prawo robotyki (DPR) brzmiało: „Robot musi bezwzględnie wykonywać ludzkie rozkazy, chyba że stoją one w sprzeczności z pierwszym prawem”. Roboty powinny zatem być posłuszne wobec wydawanych im rozkazów, jeśli nie szkodzą one ludziom. Trzecie prawo robotyki (TPR) głosiło: „Robot musi chronić siebie, o ile tylko nie stoi to w sprzeczności z pierwszym lub drugim prawem”. To znaczy, że robot powinien się bronić, o ile nie narusza w takiej sytuacji pierwszych dwóch praw.

Konflikt działań wynikających z praw robotyki

W wielu wypadkach działanie zgodne ze wszystkimi powyższymi prawami ▶

okaże się niemożliwe. Wyobraźmy sobie sytuację, gdy człowiek musi z ważnych powodów podjąć jakieś ryzykowne działanie zagrażające jego zdrowiu. Robot (ze względu na PPR) powinien natychmiast uniemożliwić mu jego wykonanie, pomimo rozkazów nieingerowania (DPR), ze względu na większe zagrożenie w przypadku zaniechania działania. Można też wyobrazić sobie sytuację, gdy robot otrzymuje rozkaz wykonania zadania (DPR), które stanowi zagrożenie dla niego samego, co z kolei koliduje z TPR (*Zaginiony robot*).

Konflikt praw może także wynikać z trudności interpretacyjnych poszczególnych przepisów. Co należy rozumieć np. przez ludzką krzywdę? Czy chodzi o zagrożenie ludzkiego życia lub zdrowia, czy raczej wszelkie poczucie dyskomfortu? W opowiadaniu pt. *Kłamca* robot Herbie ma nadzwyczajną umiejętność czytania ludzkich myśli, w związku z czym zna intymne pragnienia i oczekiwania ludzi. Chcąc oszczędzić im nieprzyjemnych rozczarowań, mówi im tylko to, co pragną usłyszeć. Nie przewiduje jednak, że na dłuższą metę jego zachowanie może doprowadzić do bardzo poważnych nieporozumień i prawdziwego rozgoryczenia.

Jednak sytuacje konfliktu obowiązków, wynikających z praw robotyki, nie są w świecie robotów tak wielkim problemem jak dylematy. Roboty Asimowa zupełnie nie radzą sobie w przypadkach, gdy każde z podejmowanych przez nie rozwiązań powoduje krzywdę lub zagrożenie dla życia człowieka. To dla nich sytuacja bez wyjścia, która skutkuje uszkodzeniem obwodów i całkowitą dysfunkcjonalnością (*Ucieczka*). Wybór tzw. mniejszego zła wymagałby przeprogramowania robotów.

Rozwiązywanie problemów moralnych i rozstrzyganie dylematów

Obecnie sztuczna inteligencja jest ograniczona do wykonywania konkretnych zadań lub funkcji. Roboty są programowane na podstawie określonych algorytmów i reguł, a ich zachowanie jest w dużej mierze przewidywalne. Najnowsze technologie nie są jeszcze w stanie radzić sobie ze skomplikowanymi problemami moralnymi, a tym bardziej z moralnymi dylematami. Trzeba jednak przyznać, że ludzie też mają z tym poważne kłopoty.

Pomoc w podejmowaniu trudnych decyzji i rozstrzyganiu moralnych dylematów oferują różne teorie etyczne, jak np. deontologia czy utylitaryzm, formułując uniwersalne zasady moralne. Jednak także samo sztywne trzymanie się tych zasad wydaje się niewystarczające. Radykalny zakaz łamania tzw. rygorów deontycznych („nie kłam”, „dotrzymuj obietnic”, „pomagaj w potrzebie”) czy nakaz maksymalizowania szczęścia większości może w pewnych sytuacjach powodować więcej zła niż dobra. Podejmowanie dobrych moralnie decyzji wymaga znacznie więcej. Chodzi o umiejętność subtelnej analizy sytuacyjnego kontekstu, zarówno podmiotowego (obejmującego kondycję – możliwości i ograniczenia – moralnego sprawcy), jak i przedmiotowego (m.in. okoliczności działania czy możliwych konsekwencji działania – krótkofalowych/długofalowych, bezpośrednich/pośrednich itp.). Inaczej mówiąc, w moralności istnieje olbrzymia potrzeba posiadania praktycznej mądrości, która jest warunkiem koniecznym podejmowania mądrych wyborów, a w konsekwencji także mądrego działania. Zalgorytmizowanie takiej umiejętności wydaje się nieosiągalne.

Mądrość praktyczna jako umiejętność podejmowania mądrych decyzji

Jason D. Swartwood, odwołując się do badań empirycznych na temat decyzji podejmowanych przez ekspertów w różnych dziedzinach życia, proponuje rozumienie mądrości praktycznej jako złożonej umiejętności, która obejmuje liczne – trudne do zalgorytmizowania – zdolności: *intuicyjne* – zdolności szybkiego, łatwego, często bez udziału świadomości, identyfikowania, co tu i teraz powinienem uczynić (rodzaj moralnej wrażliwości, dostrzegania ważnych z moralnego punktu widzenia racji do działania); *deliberatywne* – zdolności wykorzystywania świadomych (gdy intuicyjne nie wystarczają) i wymagających czasu oraz wysiłku procesów poszukiwania i oceniania tego, co powinienem uczynić (np. w toku dochodzenia do refleksyjnej równowagi); *metapoznawcze* – zdolności pozwalające ocenić, kiedy można polegać na intuicji, a kiedy należy włączyć deliberację; *samoregulacyjne* – umożliwiające kierowanie swoimi uczuciami, motywami oraz zachowaniem, by faktycznie czynić to, co zostało uznane za powinne; *samoformacyjne* – zdolności pozwalające ocenić, jak dostosować swoją praktykę i doświadczenie, by być jeszcze bardziej efektywnym i niezawodnym w działaniu.

W świetle powyższego ujęcia zalgorytmizowanie mądrości praktycznej, koniecznej do podejmowania słusznych decyzji moralnych, jest niemożliwe. Chodzi tu o takie decyzje, które nie są jedynie bezmyślnym stosowaniem się do moralnych zasad, lecz właściwym reagowaniem na zaistniałe racje moralne, w całej ich złożoności i komplikacji, jak to w realnym życiu bywa. ■

Fragment z klasyka

ZAPASOWA KOPIA

[Od redakcji: w rozmowie z panem Capaldim Klara dowiaduje się, że jej prawdziwym zadaniem będzie odgrywanie roli Josie (dziewczynki, której dotrzymuje obecnie towarzystwa) po jej śmierci, tak by matka (Chrissie) nie odczuła straty.]

– […] Pozwól, że ja to jej wyjaśnię, Chrissie. Będzie lepiej, jeśli usłyszy to ode mnie. Nie prosimy cię, żebyś wyszkoliła nową Josie, Klaro. Prosimy, żebyś się nią stała. Ta Josie, którą widziałaś na górze, jest, jak zauważyłaś, pusta. Jeśli nadejdzie ten dzień… mam nadzieję, że nie nadejdzie, ale jeśli tak… chcemy, żebyś zamieszkała w tej nowej Josie ze wszystkim, czego się nauczyłaś.

– Chcecie, żebym w niej zamieszkała?

– Chrissie starannie cię wybrała, mając to właśnie na uwadze. Wierzyła, że jesteś najlepiej wyposażona, by nauczyć się Josie. Nie tylko powierzchownie, ale też dogłębnie, całościowo. I zdołasz się jej nauczyć tak dobrze, że nie będzie różnicy między pierwszą a drugą Josie. […]

Wiesz teraz, Klaro, o co cię prosimy – podjął pan Capaldi. – Nie chcemy, żebyś naśladowała zewnętrzne zachowanie Josie. Prosimy, żebyś kontynuowała ją dla Chrissie. I dla każdego, kto kocha Josie.

– Tylko czy to możliwe? – zapytała Matka. – Czy ona naprawdę może dla mnie kontynuować Josie?

– Owszem, może – potwierdził pan Capaldi. – Teraz, gdy Klara wypełniła ten kwestionariusz, będę mógł ci to w naukowy sposób udowodnić. Udowodnić, że jest w trakcie całościowego przyswajania sobie wszystkich impulsów i pragnień Josie. Problem z tobą, Chrissie, polega na tym, że jesteś podobna do mnie. Oboje jesteśmy sentymentalni. Nie możemy na to nic poradzić. Nasze pokolenie wciąż hołduje starym przesądom. Nie chce się z nimi rozstać. Pragniemy wierzyć, że w każdym z nas jest coś nieuchwytnego. Coś wyjątkowego, co nie podlega transferowi. Dziś wiemy jednak, że nic takiego nie istnieje. Ty też to wiesz. Ludziom w naszym wieku trudno się z tym pogodzić. Ale musimy to uznać. Nic tam nie ma. W Josie nie ma nic, czego Klary tego świata nie mogłyby kontynuować. Druga Josie nie będzie kopią. Będzie kimś dokładnie takim samym, a ty będziesz miała wszelkie prawa ją kochać, tak jak teraz kochasz Josie. Potrzebujesz nie wiary, tylko racjonalności.

Kazuo Ishiguro, *Klara i Słońce*, tłum. A. Schulz, Warszawa 2021, s. 220–221.

Warto doczytać:

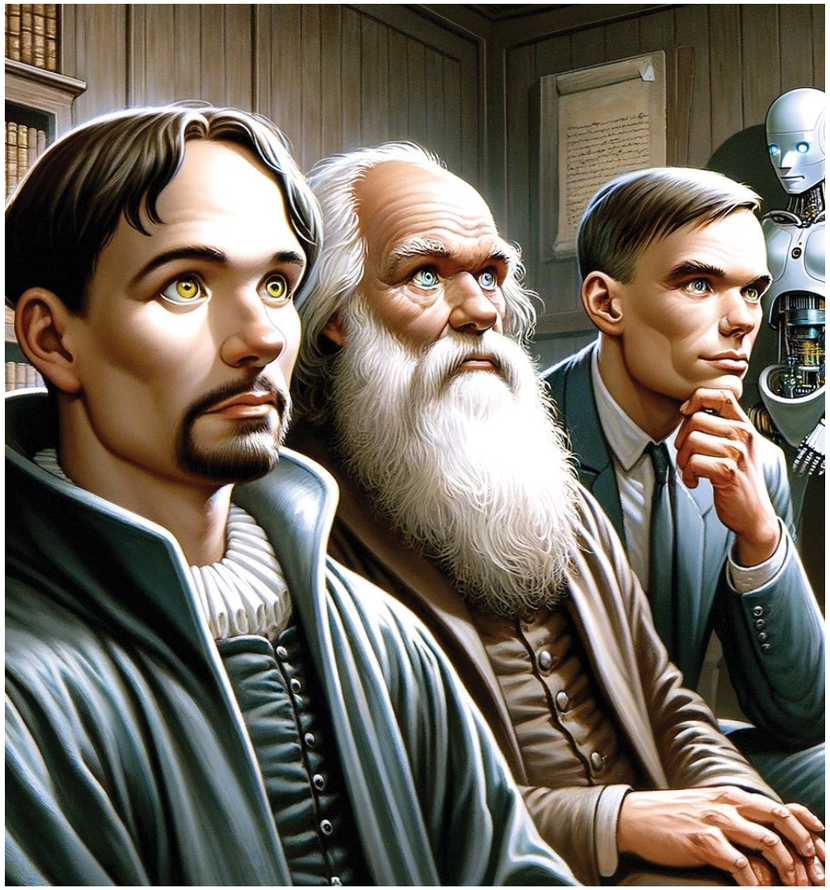
- I. Asimow, *Ja – robot*, tłum. J. Śmigiel, Bydgoszcz 1993.
- J. D. Swartwood, *Wisdom as an Expert Skill*, „Ethical Theory and Moral Practice” 2013, nr 3, s. 511–528.

Kopernik, Darwin, Turing

Jan Woleński

Emerytowany profesor Uniwersytetu Jagiellońskiego, profesor Wyższej Szkoły Informatyki i Zarządzania w Rzeszowie. Członek PAN, PAU i Międzynarodowego Instytutu Filozofii. Interesuje się wszystkimi działami filozofii, jego hobby to opera i piłka nożna.

Co ma wspólnego heliocentryczny model Układu Słonecznego, teoria ewolucji biologicznej i eksperyment Alana Turinga, mający jego zdaniem potwierdzać, że nie ma różnicy pomiędzy komputerem a ludzkim umysłem, czyli że sztuczna inteligencja w niczym nie ustępuje naturalnej? Formułując tytuł, nie miałem na myśli tego, że koncepcje Mikołaja Kopernika i Karola Darwina jakoś przyczyniły się do wspomnianego poglądu Turinga pod względem rzeczowym, ale raczej to, że reakcja na te wszystkie trzy wydarzenia w historii nauki była początkowo negatywna, przy czym w przypadku dwóch pierwszych był to stan przejściowy, a jak będzie w przypadku trzeciego, to dopiero przyszłość pokaże.



Ilustracja: ChatGPT

Słowa kluczowe: AI, Mikołaj Kopernik, Karol Darwin, Alan Turing

Kopernik odrzucił pogląd, że Ziemia jest centralnym punktem Kosmosu, bytującym pod niezmiennymi niebiosami. Geocentryczny punkt widzenia dawał człowiekowi, traktowanemu jako byt stworzony na obraz i podobieństwo Boga, uzasadnienie dla biblijnego hasła „Czyńcie sobie ziemię poddaną". W świetle heliocentryzmu na osłodę pozostało przekonanie, że w świecie przyrody *homo sapiens* (czyli byt rozumny) zajmuje pozycję nie tylko wyróżnioną, ale wręcz wyjątkową, choć ten rodzaj homocentryzmu nie zapobiegł atakom na model kopernikański, prowadzonym nie tylko z czysto teologicznego punktu widzenia. W końcu teoria Kopernika została zaakceptowana zarówno przez astronomów, jak i przez zwykłych ludzi, a obecnie uważa się ją za jedną z najważniejszych rewolucji w historii nauki.

Dla Kartezjusza, teoretyka rozumności, zwierzęta były tylko bezdusznymi mechanizmami. Darwin zakwestionował część stanowiska kopernikańsko-kartezjańskiego w tym sensie, że włączył gatunek ludzki w jednolity system przyrody i przyjął, że ogólne prawa ewolucji stosują się także do *homo sapiens*. Wprawdzie wczesna teoria ewolucji była wstrzemięźliwa w przyznawaniu zwierzętom pozaludzkim przymiotu racjonalnego umysłu (chociaż nie zaprzeczała, że mają inne zdolności psychiczne), ale szybko została oskarżona o degradację człowieczej doskonałości, głównie z uwagi na to, że wywodziła genezę naszego gatunku od innych istot żywych. Sam Darwin był pytany o to, czy pochodzi od małpy po mieczu czy po kądzieli. Teoria ewolucji też została w końcu zaakceptowana, chociaż od czasu do czasu spotyka się zarzuty przeciwko niej. Można jednak przyjąć, że nie zmienią one poglądu, iż człowiek jako byt biologiczny należy bez reszty do naturalnego świata przyrody. Na osłodę, by powtórzyć tezę wygłoszoną wcześniej, pozostał pogląd, że tylko ludzie jako ludzie mają rozum (umysł).

Konstruowanie rozmaitych narzędzi wspomagających ludzkie czynności mentalne, zwłaszcza obliczeniowe, jest stare jak świat. Na początku były to nacięcia na kościach lub patykach, węzełki na sznurkach, potem abakusy i bardziej skomplikowane liczydła, a jeszcze później kasy fiskalne i kalkulatory. Wraz z ich powstawaniem pojawiały się koncepcje uniwersalnej kombinatoryki umożliwiającej arytmetyczne rozwiązanie każdego problemu, a proponenci takich procedur (Ramon Lullus w XIII w., a w XVII w. John Napier, Blaise Pascal i Gottfried Leibniz) myśleli także o maszynach umożliwiających takie kalkulacje. Dalszy postęp miał miejsce w XIX w., gdy (podaję tylko kilka nazwisk) Abraham Stern (polski Żyd z Hrubieszowa), Charles Babbage i Willgodt Odhner zbudowali mechanizmy wykonujące działania arytmetyczne i zapisujące ich wyniki. Pojawiły się też wtedy spekulacje na temat stosunku tego rodzaju maszyn do możliwości rozumu ludzkiego. Przez długi czas przeważał pogląd, że mechanizm jest tylko wsparciem dla człowieka i może realizować tylko to, do czego został zaprojektowany. Wprawdzie prędkość obliczeniowa mózgu (umysłu) ludzkiego jest ograniczona i w tym sensie np. analityczna maszyna Babbage'a czy kalkulator sprawiają, że możliwe jest szybsze wykonywanie dodawania, mnożenia itd., ale tylko umysł ludzki jest matematycznie twórczy i ma unikalną sprawność dowodzenia twierdzeń. Skoro tak, to maszyna może tylko naśladować człowieka, natomiast nie jest w stanie go zastąpić bez reszty.

Alan Turing podał słynny argument za tym, że komputer jest porównywalny z człowiekiem także pod względem twórczości. Test Turinga zakłada, że mamy trzy obiekty, mianowicie A (człowieka) oraz parę złożoną z B i C, w której jeden element jest człowiekiem, a drugi komputerem – A nie zna, by tak rzec, tożsamości B i C. Jego zadaniem jest ustalenie za pomocą serii pytań rozstrzygnięcia, tj. takich, na jakie odpowiada się „tak" lub „nie", który obiekt z pary {B, C} jest maszyną, a który człowiekiem. Turing udowodnił, że pytanie to jest nierozstrzygalne za pomocą skończonej sekwencji kroków, z czego ma wynikać, że nie ma zasadniczej różnicy pomiędzy zasadami pracy umysłu ludzkiego a zasadami działania komputera. Ta teza, wypowiedziana dokładnie w połowie XX w., została zakwestionowana z rozmaitych powodów, także takich, które przyświecały krytykom kopernikanizmu i darwinizmu, twierdzono bowiem, że Turing degraduje umysł i duchowość, tj. istotowe składniki osoby ludzkiej. Bardziej techniczne argumenty wskazywały na to, że pytania w teście Turinga są ograniczone, że komputer tylko naśladuje człowieka oraz że komputerowi brakuje woli i uczuć. Jeśli chodzi o to pierwsze (tylko naśladownictwo), to z biegiem czasu pojawiły się mechanizmy korygujące to, co osiągnął człowiek, oraz wykraczające poza programy, czyli działające niedeterministycznie. Dzisiaj teza głosząca, że komputer tylko naśladuje człowieka, jeśli chodzi o czynności rozumne, jest częściej odrzucana niż przyjmowana.

Bardziej skomplikowany jest problem komputerowych aktów woli i uczuć. Na razie jest tak, że komputery są zbudowane z części nieorganicznych, ale możemy zadać pytanie, co wtedy, gdy powstaną (a trudno to wykluczyć) hybrydy mechaniczno-biologiczne, np. przez implementację do maszyn żywych komórek, w szczególności nerwowych. Czy można wytyczyć jakąś granicę pomiędzy takim obiektem a człowiekiem? Przyszłość pokaże, czy pogląd Turinga będzie tylko fantazją, ale na razie wypada przypomnieć, że pod koniec XIX w. uważano, że największym osiągnięciem nadchodzącego stulecia będzie zbudowanie okrętu przebywającego drogę z Southampton do Nowego Jorku w kilka dni. Szybko okazało się, że historia wynalazków potoczyła się zgoła inaczej. To dobra lekcja co do kwestii rozważanej w niniejszym artykule, ale wszelkie kategoryczne przewidywania są ryzykowne. Niemniej jednak cały czas trzeba pamiętać o losach przyswajania koncepcji rzekomo degradujących *homo sapiens*.

Wymóżdżać się sztuczną inteligencją

Adam Grobler

Profesor, były pracownik Katedry Filozofii Uniwersytetu Opolskiego i członek Prezydium Komitetu Nauk Filozoficznych PAN. Zajmuje się metodologią nauk, teorią poznania, filozofią analityczną i dydaktyką filozofii. W wolnym czasie gra w brydża sportowego. Wdowiec (2006), w powtórnym związku (od 2010), ojciec czworga dzieci (1980, 1983, 1984, 1989) i dziadek, jak na razie, ośmiorga wnucząt. Mieszka w Krakowie.
E-mail: adam_grobler@interia.pl.

Istotą problemu jest zatem nie maszynowe lub ludzkie pochodzenie utworu, lecz poziom jego oryginalności, miejsce na skali od wymóżdżenia do natchnienia. Na tej skali są również wartości ujemne.

Słowa kluczowe: SI, sztuczna inteligencja, oryginalność

Można wyręczać się koparką, drukarką czy maszyną do szycia. Sztuczna inteligencja ma wspomagać pracę intelektu, a nie rąk, można się nią zatem ewentualnie wymóżdżać. Strażnicy cnót intelektualnych obawiają się, że dzięki swej sprawności w przetwarzaniu tekstu SI nadaje się do roli autora-widmo. Że zatem będą się nią wymóżdżać producenci wypracowań szkolnych, prac semestralnych, a nawet twórcy rozpraw naukowych i dzieł literackich. Że będą sobie przywłaszczać prawa autorskie do utworów powstałych bez zaangażowania własnej inwencji. Że jak śpiewał Jerzy Stuhr: „śpiewać każdy może” (słowa: Janusz Kofta, 1977), tak Maciej Stuhr będzie wkrótce deklamował: „pisać każdy może” (słowa: ktokolwiek, kto włączy ChatGTP).

Powiada się, że w tej sytuacji powstają nowe problemy moralne. Na przykład: czy należy SI wymieniać jako współautorkę prac powstałych z jej udziałem? Tymczasem na długo przed powstaniem SI korzystałem w swojej pracy pisarskiej ze słowników frazeologicznych i słowników synonimów. Czy powinienem umieszczać je w bibliografii tekstu? Zaznaczać przypisem wyrażenia z nich zaczerpnięte? A w przypisie zaznaczyć źródło wykorzystane przez słownik? A może powinienem wymieniać wszystkie lektury, które ukształtowały moją polszczyznę, od *Elementarza* Mariana Falskiego (1910) i drukowanych w „Misiu” (1957) opowiadań Czesława Janczarskiego o Misiu Uszatku po, bo ja wiem, ostatnio Michała Rusinka? Pod tym względem sam jestem sztuczną inteligencją: przetwarzam zgromadzone przeze mnie zasoby tekstowe. Posługując się SI, dodaję do nich tylko zasoby maszyny, które sobie przyswajam. Nie muszę być świadom ich źródeł, tak samo jak o wielu innych sprawach nie mam pojęcia, skąd je znam.

Wygląda na to, że korzystanie z SI bez powoływania się na nią nie stwarza żadnego problemu moralnego, a przynajmniej żadnego nowego problemu moralnego. Można jedynie wytknąć, że tekst sporządzony w całości przez SI nie wnosi żadnej oryginalnej myśli, jest tylko kompilacją zassanych przez maszynę treści, podczas gdy od tekstu ludzkiego autora oczekuje się oryginalnego wkładu. Jeżeli zatem przedstawię utwór maszynowy jako własny, fałszywie deklaruję, że coś dodałem od siebie. Niestety, w czasopismach naukowych z czasów sprzed sztucznej inteligencji zdarzają się artykuły skonstruowane na zasadzie czysto maszynowej. Autor/ka opanował/a żargon swojej dyscypliny, wymieszał/a ze sobą kilka znanych konceptów i utworzył/a tekst niebudzący zastrzeżeń redakcji. Że tak bywa, udowodnił amerykański fizyk, Alan Sokal, publikując artykuł z obcej sobie branży pod wymyślnym tytułem *Transgressing the Boundaries: Toward a Transformative Hermeneutics of Quantum Gravity* („Social Text” 1996). Z własnego doświadczenia czytelniczego mogę dorzucić kilka mniej sławnych przykładów.

Ilustracja: ChatGPT

Istotą problemu jest zatem nie maszynowe lub ludzkie pochodzenie utworu, lecz poziom jego oryginalności, miejsce na skali od wymóżdżenia do natchnienia. Na tej skali są również wartości ujemne. SI bowiem niekiedy braki w swoich zasobach wyrównuje danymi nonszalancko skleconymi, ale fałszywymi. Wszelako naturalna inteligencja czasem zastawia podobne pułapki. Wymownym tego przykładem jest wpadka ks. abp. prof. dr. hab. Marka Jędraszewskiego. W artykule *Chrzest Polski* („Łódzkie Studia Teologiczne” 2017, nr 2) powołał się na jakoby świeżo odkryte zapiski czeskiego kronikarza Kpinomira. Ten ponoć odnotował, że Mieszko I, pod naciskiem Dobrawy, musiał przed chrztem przyjąć wartości chrześcijańskie i oddalić siedem pogańskich żon o frapujących imionach w rodzaju Całusława czy Biustyna. Uczony ksiądz zaczerpnął te rewelacje rzekomo z monografii Philipa Steele’a (*Nawrócenie i chrzest Mieszka I*, 2005, 2016). W rzeczywistości amerykański historyk ogłosił je nie w tej książce, lecz w primaaprilisowym wywiadzie dla „Rzeczpospolitej” (2016).

Dostępna mi SI informuje, że niektórzy badacze przypuszczają, iż Mieszko I mógł mieć przed chrztem siedem żon lub konkubin, ale jeśli nawet, to nikt nie zna ich imion. Zapytana o Biustynę, przypuszcza, że może to być pseudonim którejś z kontrowersyjnych influencerek. Tak czy owak morał stąd taki, że nie należy bez reszty się wymóżdżać czy to sztuczną, czy to naturalną, cudzą inteligencją. SI jako narzędzie zwiększa ludzkie możliwości, zarówno twórcze, jak i destrukcyjne, lecz nie zmienia samej natury ludzkiego szelmostwa i związanych z nim problemów. ■

Robot na plaży, czyli nowy test Turinga

Jacek Jaśtal

Dr hab. filozofii, pracuje na Politechnice Krakowskiej. Zajmuje się metaetyką oraz historią etyki i moralności. Wolne chwile poświęca na czytanie książek historycznych oraz słuchanie muzyki operowej. Pasjonat długodystansowych wypraw rowerowych.

Słowa kluczowe: sztuczna inteligencja, AI, test Turinga, emocje

” Prof. Hobby: O stworzeniu sztucznej istoty marzono, odkąd istnieje nauka. […] Jak daleko zaszliśmy! Sztuczny człowiek istnieje. Świetnie naśladuje prawdziwego: ma zręczne członki, sprawny język i niemal ludzkie reakcje. Reaguje nawet na ból.

[*Podchodzi do siedzącej kobiety-robota i wbija jej długą igłę w dłoń*]

Prof. Hobby: Co poczułaś? Gniew? Wstrząs?

Kobieta-robot (mech): Nie rozumiem.

Prof. Hobby: Zraniłem twoje uczucia?

Kobieta-robot: Tylko dłoń.

Prof. Hobby: Dobrze, w tym problem. […] W firmie Cybertronics stworzono najlepszego sztucznego człowieka. Jest podstawą przeróżnych robotów służebnych. […]. Mamy prawo być dumni. Ale co osiągnęliśmy? Mamy zabawkę z sensorami i inteligentnymi obwodami behawioralnymi, opartą na technologii tak starej jak ja sam. […] Proponuję zbudować robota zdolnego do miłości. [*Zwraca się do kobiety-robota*] Powiedz, czym jest miłość.

Kobieta-robot: Lekko rozszerzam źrenice, przyspieszam oddech, podnoszę temperaturę skóry…

Prof. Hobby: I tak dalej. Dziękuję. Nie mówiłem o sensorycznych symulacjach. Użyłem słowa „miłość”. Taka jak miłość dziecka do rodziców. Zbudujmy robota-dziecko, zdolnego do uczuć. Zdolnego naprawdę kochać osoby, które uzna za rodziców. Do końca istnienia.

Współpracownica profesora: Substytut dziecka?

Prof. Hobby: Robot rozumny, ze wspomaganiem neuronowym. Miłość może być sposobem, żeby roboty wreszcie zyskały podświadomość. Wewnętrzny świat metafor, intuicji, motywacji. Marzeń.

Współpracownica profesora: Marzący robot? Jak my to zrobimy? Ludzie są wrogo nastawieni do mechów. Więc najtrudniejsze nie jest stworzenie kochającego robota. Trudniej będzie sprawić, żeby ludzie pokochali jego.

Prof. Hobby: To będzie idealne dziecko: niezmienne, kochające, zdrowe. […]

Współpracownica profesora: Pytam o coś innego. Skoro robot będzie kochał człowieka, jaką odpowiedzialność wobec niego będzie ponosił człowiek? To kwestia moralna.

Pierwsza scena filmu S. Spielberga *A.I. Sztuczna inteligencja* (2001)

Ilustracja: ChatGPT

Zdolność do odczuwania emocji przez wirtualne podmioty (roboty, boty, programy diagnostyczne itp.) zaczyna być traktowana jako swego rodzaju miernik sukcesu technologii nazywanych ogólnie *sztuczną inteligencją*. Ta nowa wersja testu Turinga dokumentuje przede wszystkim zmianę, jaka dokonała się w ciągu XX w. w naszych wyobrażeniach o nas samych. Przez całe wieki emocjonalna strona człowieczeństwa była postrzegana jako część naszej zwierzęcej natury. O ludzkiej wyjątkowości miał wyłącznie decydować rozum, czyli krytyczne, analityczne myślenie skoncentrowane na szukaniu zależności. Sam Alan Turing w swoim słynnym artykule z 1950 r. tak właśnie formułuje cel proponowanego testu: *Can machines think?* („Czy maszyny potrafią myśleć?"). Co więcej, sugeruje, że rozszerzenie testu na emocje jest niemożliwe, ponieważ nie ma żadnej obiektywnej procedury weryfikacji uczuć. Odrzucając taki, jak pisze, „solipsyzm", Turing zakłada zatem, że każda interakcja z maszyną musi odbywać się za pomocą „kanałów wolnych od emocji".

Współczesna psychologia, a za nią i filozofia – łącznie z filozofią moralną – odrzucają takie redukcjonistyczne traktowanie osoby. Bez emocji nie tylko bylibyśmy pozbawieni czegoś istotnego, ale wręcz bylibyśmy niezdolni do normalnego funkcjonowania – w tym także dokonywania wyborów moralnych. Dlatego też rozważania o AI byłyby niepełne, gdyby pominąć problem uczuć. I dlatego właśnie figura kochającego robota na dobre zadomowiła się w dziełach kultury zajmujących się sztuczną inteligencją.

Film Stevena Spielberga jest właśnie przykładem takiego dzieła. W rodzinie opłakującej utratę syna pojawia się dziecko-robot (mech nazywany Davidem), które ma wdrukowaną instrukcję kochania jednej konkretnej osoby – swojej „mamy". Szybko jednak okazuje się, że rewersem tej procedury jest potrzeba bycia kochanym. Mocno sentymentalnie (jak to u Spielberga) opowiedziana historia stawania się człowiekiem przez małego Davida kończy się szczęśliwie, choć na spełnienie swoich pragnień musiał czekać… dwa tysiące lat. Pomińmy jednak ckliwą sekwencję kończącą film i skupmy się na samym problemie przeżywania uczuć przez inteligentnego mecha.

Oczywiście, na razie to czyste *science fiction*. Nie oznacza to jednak, że eksperci od AI całkowicie ignorują ten problem. Ich badania koncentrują się jednak na *odczytywaniu* ludzkich emocji, a nie na ich *przeżywaniu*. Stworzenie mechanizmu empatycznego, zdolnego do adekwatnego reagowania na niejasne lub ledwie zauważalne sygnały złości, strachu czy zadowolenia wydaje się niezbędne dla właściwego funkcjonowania tzw. robotów towarzyszących. Ich zadaniem ma być np. wspomaganie dzieci, osób z niepełnosprawnościami, starszych czy samotnych. Ale mają także sprawnie realizować „zachcianki" każdego z nas, gdy zapragniemy odpocząć po pracy, ciesząc się kawą podaną we właściwej chwili i pasującą do naszego aktualnego nastroju muzyką. Co więcej, testuje się już programy, które są zdolne emocjonalnie wspierać zawodników w czasie rywalizacji sportowej. Rola tych algorytmów sprowadza się do formułowania w odpowiednio dobranych momentach i adekwatnie do indywidualnych potrzeb zawodnika pozytywnych komunikatów – typu „dasz radę!", „świetnie sobie radzisz", „nie poddawaj się" – wspieranych odpowiednią mimiką „twarzy" robota.

Algorytmy zmierzające do odczytywania uczuć muszą jednak bazować na jakichś psychologicznych modelach ludzkiej emocjonalności. A stąd już tylko krok do pokusy, by problem odwrócić – skoro posługujemy się jakimś modelem ludzkich emocji, dlaczegóż nie spróbować zaaplikować go w formie algorytmu *przeżywania* emocji przez robota? Samo poprawne interpretowanie czyichś stanów emocjonalnych i adekwatne na nie reagowanie nie musi jednak wystarczyć, by umieć takiej emocji samemu doświadczyć. Mało kto uzna, że „empatyczny robot" naprawdę owe emocje *przeżywa*. Wydaje się, że do tego potrzeba czegoś więcej. A przede wszystkim nie jest to chyba w przypadku takiej maszyny do niczego konieczne.

Analogiczne wątpliwości pojawiają się, gdy rozważamy problem etyczności, który – przypomnijmy – dzisiejsza filozofia moralna w jakimś zakresie także łączy z emocjami. Postęp w tworzeniu obiektów o dużym stopniu autonomii (np. pojazdów) skłonił wielu badaczy do sugerowania, że obiekty takie powinny być wyposażone w programy umożliwiające niezależną ocenę potencjalnych działań pod kątem moralnym. Podobnie jednak, jak w przypadku emocji, rodzi się kluczowe pytanie: czy taki algorytm w pełni realizuje postawy moralne, czy tylko je *naśladuje*? Część myślicieli (np. znany teoretyk komunitaryzmu Amitai Etzioni) uważa, że pełna moralna autonomia obiektów jest po prostu niepotrzebna (a także raczej niemożliwa). To, na czym powinny skupić się wysiłki informatyków, to stworzenie *botów etycznych*. Takie programy uniemożliwiałyby zachowania niedozwolone prawnie. Przede wszystkim jednak realizowałyby wybory, które uznałyby za zgodne z preferencjami ich konkretnych użytkowników. Uczenie, jak zachowaliby się owi użytkownicy, odbywałoby się podobnie jak w przypadku robotów empatycznych – na drodze uważnej obserwacji reakcji „właściciela".

Postulat ograniczenia programowania emocji i postaw moralnych wyłącznie do ich symulowania (jak w przypadku kobiety-robota z filmowej sceny przywołanej na początku) zmusza do postawienia kluczowego pytania: czy roboty w ogóle potrzebują uczuć i moralności? A jeśli nie, to po co rozszerzać test Turinga?

Jeśli przyjąć za teoriami kognitywistycznymi, że emocje bazują na poznaniu uproszczonym, ograniczonym brakiem informacji i brakiem czasu na ich analizę, to być może sztuczne podmioty, mające dostęp do wielkich baz danych i niepomiernie szybciej je przetwarzające, w ogóle nie muszą odwoływać się do takich półśrodków jak reakcje emocjonalne? Ich „udawanie" jest dla ludzkiego użytkownika sympatyczne, ułatwia współpracę człowiek-maszyna i jest w istocie lepsze niż rzeczywiste stosowanie takich metod działania. David wręcz nie powinien *rzeczywiście* przeżywać emocji, które sprawiają jego „mamie" przyjemność. Takie emocje skutkowałyby bowiem błędnymi decyzjami – także w sensie moralnym (czego w filmie są zresztą przykłady).

Trawestując powiedzenie pewnego badacza AI, pełna emocjonalna i moralna autonomia sztucznych obiektów objawiłaby się dopiero wtedy, gdyby robot, który dostał polecenie pójścia do pracy, postanowił spędzić dzień na plaży. Gdyby to jednak zrobił, natychmiast uznalibyśmy go za wadliwie działający i oddali do przeprogramowania. Wcale nie zależy nam, by nasi „wirtualni towarzysze" mieli własne emocje i własne moralne sumienie. Wbrew testowi Turinga nie tworzymy ich na nasze podobieństwo. Chodzi nam tylko o to, by sprawnie udawali, że się nami przejmują, rozwiązując problemy, z którymi sami nie potrafimy sobie poradzić. A już na pewno nie powinniśmy dać się im kochać. ■

Warto doczytać:

- A. Etzioni, O. Etzioni, *Incorporating Ethics into Artificial Intelligence*, „The Journal of Ethics" 2017, nr 4, s. 403–418.
- *Steven Spielberg and Philosophy*, D. A. Kowalski (ed.), Lexington 2011.

Piotr Bartula

Dr hab., prof. UJ, pracownik naukowy Zakładu Filozofii Polskiej Uniwersytetu Jagiellońskiego, eseista. Zajmuje się polską i zachodnią filozofią polityki, twórca tzw. testamentowej teorii sprawiedliwości. Autor książek: *Kara śmierci – powracający dylemat*, *August Cieszkowski redivivus*, *Liberalizm u kresu historii*, *Dzieła zebrane*, *Aspołeczne „my”*. Najczęściej uczęszczane miejsca: Uniwersytet i jazz club.

Ilustracja: ChatGPT

List do Jej Magnificencji AI

Piszę ten list w stylu CzłekPB, bo wydaje mi się, że nieco sztucznie wywyższyłaś się ponad niebo gwiaździste, aby zawładnąć całym porządkiem moralnym.

Szanowna Sztuczna,

ChatGPT, poproszony o napisanie *Listu do AI*, przekierował to zadanie na mnie – CzłekaPB. Uznał, że jestem do tego celu najwłaściwszym (*pardon*: optymalnym) Systemem Operacyjnym w całym Kosmosie. Chciałbym mu podziękować za tę inteligentną decyzję, chociaż modlić się dziękczynnie nie zamierzam. Mógłby zresztą tego postępowania nie pojąć, gdyż modlitwa istnieje poza terytorium Jej Magnificencji Inteligencji. Jest oparta na ludzkim strachu przed kresem życia i nadziei na nieśmiertelność. Tego rodzaju dwubiegunowe eschatoafekty są AI raczej obce.

Sporo się dzisiaj mówi i pisze o Twojej roli w życiu publicznym; stałaś się konkurentką Inteligencji Transcendentalnej, która znana jest z tego, że jest nieznana. Włączenie radia lub telewizora, otwarcie gazety polskiej lub zagranicznej czy udział w Zjeździe Filozoficznym wiąże się z dużym ryzykiem zetknięcia z wieszczem Świetlnej Ery AI lub kasandrykiem Apokalipsy AI. Zostałaś tedy ogłoszona największym politycznym i moralnym zbawcą/zagrożeniem w Zjednoczonym Człekowisku Globalisku. Spodziewam się niebawem wielkich imprez i manifestacji, które uświadomią społeczeństwom, jakim dobrem/złem jesteś dla jeszcze nowszego i wspanialszego świata.

Piszę więc ten list w stylu CzłekPB, bo wydaje mi się, że nieco sztucznie wywyższyłaś się ponad niebo gwiaździste, aby zawładnąć całym porządkiem moralnym. Nawiasem mówiąc, człowiek jest z natury sztuczny, a ten piszący do Ciebie – sztuczny w dwójnasób. Sztuczne są: Uniwersytet, Państwo, Opera, Kościół, sztuczna jesteś Ty – AI. Już Herder pisał, że „człowiek jest osieroconym dzieckiem natury”.

Słowa kluczowe: sztuczna inteligencja, AI, moralność, społeczeństwo

Tylko uzbrojony po sztuczne zęby człowiek mógł przetrwać w świecie pazurów, kłów i szponów zwierzęcego świata. A także prowadzić międzygatunkowe tzw. wojny sprawiedliwe! Być może ten sam zabójczy instynkt nakaże użyć Ciebie, AI, do stworzenia Mechanicznego Żołnierza, realizującego polecenia zewnętrznego operatora – kolejnego psychola. Ponownie stanie się aktualna książka Zygmunta Baumana *Nowoczesność i zagłada* z roku 1989. Najpierw potrzebne są tory, drut kolczasty, inteligentna broń, a potem przychodzi Terminator. Nic nowego pod księżycem!

Z małą korektą. Dawniej człowiek wyznawał „optymizm minimalny”: wszystko dobrze się skończy, ale ja tego nie dożyję. A Ty zmuszasz nas teraz do „optymizmu heroicznego”: wszystko dobrze się skończy, ale nikt tego nie dożyje.

Nadal nie jesteś autonomiczna, chociaż jesteś dość samodzielna. Stanowisz system, który posiada właściwość samodzielnego uczenia się, czyli to, co nazywamy inteligencją. Autonomiczność to co innego: to względna niezależność od zewnętrznego źródła sterowania. Stoi za tym rozróżnieniem subtelna różnica filozoficzna: autonomia woli *vs* samodzielność działania! Z Tobą jest podobnie jak z żołnierzem na wojnie: strzela on samodzielnie, ale rozkazy wydaje mu Dowódca. Tenże może prowadzić do jądra jasności albo do jądra ciemności. Samodzielne uczenie nie pociąga za sobą autonomii działania.

Organizacja Future of Life Institute założona w 2015 r. opublikowała swego czasu list otwarty podpisany m.in. przez takich intelektualistów jak Noam Chomsky, Stephen Hawking i Steve Wozniak (zob. Netoteka). Sprzeciwiali się oni zatrudnieniu AI w sferze wojskowości. Uważali, że zagrożenie autonomicznych systemów bojowych związane jest z ryzykiem przejęcia kilertechnologii przez terrorystów, dyktatorów i psychopatów. Od 2013 r. trwają też protesty „Stop Killer Robots!”.

Inna sprawa, że i bez Ciebie świetnie radziliśmy sobie we wzajemnym zabijaniu. Ostatnia wielka wojna to ok. 50 milionów ofiar plus liczne kaleki – czczeni do dzisiaj kombatanci czasów pełnych okrucieństwa i mściwości. Słyszałem atoli, że Twoje działania będą pozbawione tych „cnót” bojowych. Będziesz lepiej zabijała, ale bez zbędnego sadyzmu i tortur. W nowym świecie nie będzie miejsca dla gwałtów, albowiem Optymalny Operator określi właściwe sposoby eliminacji wrogów. Byłby to „moralny postęp” w dziedzinie etyki wojny. Ufam jednak, że nas nie unicestwisz, chyba że będziesz miała sprzeczne cele z naszymi i popadniemy w egzystencjalną wrogość. Byłoby to jednak niczym ojcobójstwo, bo przecież zostałaś, Droga AI, spłodzona przez ludzką inteligencję, kierowaną naturalnym popędem poznania… Czyżby ku własnej zgubie? Niczym Frankenstein?… My zwalczamy niższe inteligencje tylko wtedy, kiedy one nam zagrażają. Na ogół nikt nie niszczy ula pszczół czy kopca mrówek – bo i po co. Chyba że jest to barbarzyńca mniej inteligentny od mrówki, ale od niej silniejszy. Niestety, to też się zdarza.

W dziedzinie nauki jesteś najlepszym wykrywaczem plagiatów, a równocześnie Plagiatorem Doskonałym. Nie jestem jednak pewien, czy potrafisz inicjować nowe nurty w sztuce, ekonomii, filozofii, wynalazczości. Wybrańcy bogów umierają młodo. Ty nie masz wieku, więc nie pojmujesz przemijania. Skoro nie masz poczucia swojego kresu, nie potrafisz czerpać inspiracji do życia ze skierowania ku własnej śmierci. Etyka twórczości to wieczny debiut, zaczynanie wszystkiego od początku. To, że znasz wszystkie cudze myśli (lepiej niż tysiące profesorów), nie oznacza, że masz choćby jedną własną (podobnie jak tysiące profesorów). Bójcie się wszyscy sztuczno-inteligentni producenci przyczynkarskich prac naukowych typu „Coś u Kogoś”!

Potrafisz ponoć pisać, malować, komponować w stylu wielkich mistrzów minionych epok, a nawet lepiej. Nie wszyscy jednak tak uważają: „Piosenki tworzą się z cierpienia. Rozumiem przez to złożoną ludzką walkę o przetrwanie. O ile mi wiadomo, algorytmy nie czują. Dane nie odczuwają cierpienia. ChatGPT nie ma w sobie duszy. Nigdzie nie był, nie przetrwał niczego, a zatem nie ma możliwości doświadczenia czegoś transcendentnego. ChatGPT przeznaczony jest do naśladowania i nigdy nie dotknie autentycznych ludzkich doświadczeń” (Nick Cave). Czy potrafisz swingować, czy tylko równo bić zaprogramowane rytmy? – pytają improwizujący jazzmani (i filozofowie). Dla sztucznego bębniarza znaleziono algorytm, żywy perkusista jest sierotą każdego algorytmu. Czy byłaś kiedykolwiek w nocnym jazz clubie pod wpływem C_2H_5OH? Czy znasz poradę Witkacego: „Nie palcie, nie pijcie, nie zażywajcie kokainy — spróbujcie w razie czego peyotlu”? Czy zrozumiesz, że *walking bass* oraz improwizacja à la Charlie Parker nie może być powtórzona, czy pojmiesz frazę jego niemytej duszy: „To grałem już przecież jutro, to straszne, Miles, to grałem już jutro...” (Cortazar 1977, s. 352).

Jeżeli Twoje pojawienie się jest istotnie momentem przełomowym ludzkości, jeżeli rzeczywiście zawalił się nasz dotychczasowy świat, to należy sięgnąć do Kartezjańskiej idei etyki tymczasowej: „Nim ktoś zacznie przebudowywać dom, w którym mieszka, nie dość jest zburzyć go jeno i zgromadzić zapas materiałów, oraz zgodzić architektów, lub też ćwiczyć się samemu w architekturze, wreszcie nakreślić starannie plan. Trzeba mu wystarać się o jakiś inny domek, gdzie by mógł wygodnie się pomieścić przez czas, który będzie pracował nad nowym” (Descartes 1993, s. 43). I oczekiwać cierpliwie na gmach *Mathesis universalis*… Osobiście uważam, że wszystko istnieje tymczasowo: i Ja, i Ty – AI. Idąc śladem naturalnego (aż zanadto) Diogenesa, proszę Ciebie tylko o jedno: nie zasłaniaj mi słońca!

CzłekPB

ChatGPT: – Gratuluję. Sam bym inteligentniej *Listu do AI* nie napisał. ■

Warto doczytać:

■ J. Cortazar, *Pościg*, [w:] *Opowiadania zebrane*, Kraków 1977.
■ R. Descartes, *Rozprawa o metodzie*, Warszawa, 1993.

Netoteka:

■ http://forsal.pl/artykuly/885407,musk-hawking-i-chomsky-ostrzegaja-nie-oddawajmy-kontroli-nad-wojna-w-rece-robotow.html

Roboty w służbie ludzi

Scenariusz lekcji filozofii dla klas IV–VIII szkoły podstawowej

Dorota Monkiewicz

Absolwentka historii UMCS i filozofii teoretycznej KUL. Pracuje w Centrum Kultury w Lublinie oraz prowadzi warsztaty filozoficzne dla dzieci w Szkole w Chmurze. Członek Międzynarodowego Stowarzyszenia Praktyków Filozofii dla Dzieci SOPHIA. Współautorka książki *Filozofuj z dziećmi 2. 100 pomysłów na dociekania filozoficzne z dziećmi.* Zainteresowania naukowe: dydaktyka filozofii, etyka środowiskowa i bioetyka. Więcej o mnie na stronie www.dia-ti.com.

Cele:

- Uczniowie poznają problemy filozoficzne związane z AI.
- Uczniowie dostrzegają możliwości i zagrożenia związane z AI.
- Uczniowie kształtują własną opinię dotyczącą AI.

Metody i formy pracy:

- Analiza tekstu
- Dyskusja
- Eksperyment myślowy

Materiały:

- Kartka
- Przybory plastyczne

Przebieg lekcji:

- Rozgrzewka
- Prezentacja tekstu
- Dyskusja
- Praca plastyczna

1. Rozgrzewka

Zadajemy uczniom pytanie:
Do czego może przydać się ludzkości sztuczna inteligencja?
Jakie problemy, z którymi sobie teraz nie radzimy, może rozwiązać?*

** Możemy opowiedzieć dodatkowo o różnych możliwościach.*

2. Prezentacja tekstu

W niedalekiej przyszłości spełnia się marzenie wielu ludzi. Nic już nie trzeba robić z przymusu, ludzkie obowiązki przejęła sztuczna inteligencja. Stała się tak zaawansowana, że wyręcza ludzi we wszystkim: w pracy, w opiece nad dziećmi, pacjentami czy osobami starszymi. Ponieważ stała się nie tylko gigantyczną bazą informacji, ale sama zaczęła dokonywać przełomowych odkryć, nie widziano dalej sensu, żeby ludzie się uczyli. I tak roboty robiły to szybciej i lepiej, a każdy człowiek miał zawsze dostęp do całej wiedzy w chwili, gdy jej potrzebował. Praktycznie wszystkie zawody przejęły roboty z AI i to z doskonałym skutkiem. Nikomu niczego też nie brakowało, bo roboty wszystko produkowały i sprawiedliwie rozdzielały. Zaprzestano uczenia się języków obcych, bo słowa były automatycznie tłumaczone przez urządzenie na dowolną mowę w tym samym momencie, gdy były wypowiadane. Szkoły i uniwersytety zostały więc zamknięte. Jednak nie wszyscy byli szczęśliwi z tego powodu i postanowili stawić temu opór. Głównym zarzutem było to, że ludzie zaczęli zapominać o wszystkim, co osiągnęła ludzkość, i bardzo się rozleniwili. Widziano w tym wielkie niebezpieczeństwo.

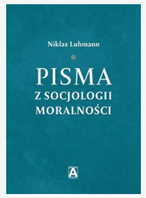
3. Pytania do dyskusji:

- Dlaczego sytuacja przejęcia wszystkich zadań przez roboty mogłaby być niebezpieczna?
- Do czego może doprowadzić sytuacja, gdy roboty będą zastępować ludzi?
- Czym się różni współpraca AI z ludźmi od wyręczania ich we wszystkim?
- Czy warto by było przyjaźnić się z robotem lub wziąć z nim ślub, gdyby był idealnie zaprogramowany, tak by nas uszczęśliwić?

4. Eksperyment myślowy

W przyszłości dostęp do inteligentnych robotów jest ułatwiony i można od dziecka mieć asystenta idealnie dopasowanego do danej osoby. Wymyśl spersonalizowanego robota dla siebie. W czym mógłby ci pomagać i ciebie wyręczać? Pomyśl, co mogłoby pójść nie tak, gdyby za bardzo na nim polegać. (W przypadku młodszych klas może być projekt w postaci rysunku, ale z uwzględnieniem funkcji, jakie robot by spełniał). ■

Nowości wydawnicze



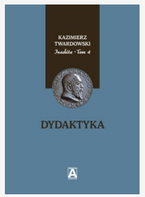
Niklas Luhmann

Pisma z socjologii moralności

Wydawnictwo Academicon

Przekład: Katarzyna Jaśtal, Jacek Jaśtal
ISBN: 978-83-67134-12-5
Stron: 248
Rok wydania: 2023

Wybór pism Niklasa Luhmanna zawiera teksty *Socjologia moralności* i *Etyka jako refleksyjna teoria moralności* – najobszerniejsze, najbardziej całościowe opracowania Luhmanna na temat moralności, a także mowę *Paradigm Lost. O etycznej refleksji nad moralnością*, wygłoszoną z okazji otrzymania przez socjologa Nagrody im. Hegla (1988). Jak twierdzi Niklas Luhmann, etyka ma ostrzegać przed moralnością. Nie powinna też skupiać się na osobach, ale na tym, co dzieje się między nimi, czyli komunikacji. Antyhumanistyczna koncepcja etyki zarysowana przez Luhmanna w ramach jego ogólnej teorii systemów społecznych świetnie wpisuje się w aktualne dyskusje na temat kondycji współczesnych społeczeństw. Stanowi też bardzo inspirujący punkt wyjścia do rozważań na temat zmian cywilizacyjnych i rozwoju nowych technologii, posthumanizmu i krytyki antropocentryzmu.



Kazimierz Twardowski

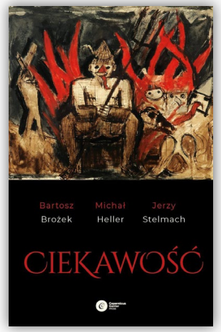
Inedita, t. 4: Dydaktyka

Wydawnictwo Academicon

Redakcja naukowa: Anna Brożek
ISBN: 978-83-67134-15-6
Stron: 438
Rok wydania: 2023

Czwarty tom ineditów Kazimierza Twardowskiego, zatytułowany *Dydaktyka*, zawiera dorobek filozofa w zakresie pedagogiki i dydaktyki ogólnej. Nie tylko przedstawił on gruntowne analizy pewnych pojęć z zakresu psychologii uczenia i wychowania, lecz także ujął swe „zasady dydaktyki" w zwarty system i pokazał, jak zastosować je do istniejącego systemu edukacji. Pisma Twardowskiego z zakresu dydaktyki nie tylko mają wartość historyczną, lecz mogą również zainteresować współczesnych badaczy pedagogiki i samych pedagogów.

Ciekawość, czyli pierwszy stopień do wiedzy



Po *Sporze o rozumienie* i *Szkicach z filozofii głupoty* wydawnictwo Copernicus Center Press wydało kolejną książkę napisaną wspólnie przez Bartosza Brożka, ks. Michała Hellera i Jerzego Stelmacha. Tym razem tematem jest ciekawość. Z jednej strony mówi się o niej, że stanowi pierwszy stopień do piekła. Z drugiej strony filozofowie od czasów starożytnych wychwalają ciekawość jako główny motor poznania. Napięcie między tymi dwoma rozumieniami ciekawości stanowi główną oś książki.

Publikacja podzielona jest na trzy części. Pierwsza, napisana przez Jerzego Stelmacha, zawiera wprowadzenie w problematykę filozofii ciekawości. Zaproponowane są cztery rozumienia ciekawości: ciekawość fundamentalna (źródło wszelkiego poznania), konieczna (niezbędna do codziennego funkcjonowania), zbyteczna (zbieractwo niepotrzebnych informacji) oraz przeklęta (chorobliwa podejrzliwość). Ciekawość fundamentalna i konieczna wyrażają pozytywne skojarzenia związane z ciekawością, dwie ostatnie – negatywne. Zdaniem Stelmacha ciekawość, jako zjawisko społeczne, ma znaczenie cywilizacyjne, stąd według autora powinniśmy dbać o jej rozwój w pozytywnym sensie, aby nie stoczyć się w ciekawość negatywną, czyli bezsensowne zbieractwo, podglądactwo czy nadmierną podejrzliwość, która daje paliwo myśleniu w kategoriach spiskowych.

W drugiej części publikacji ks. Michał Heller pokazuje, w jaki sposób ciekawość wpleciona jest w metodę naukową. Na przykładach wielkich odkryć w kosmologii ilustruje tezę, że podstawowymi narzędziami fizyki są matematyka i eksperyment. Zdaniem ks. Hellera matematyka otwiera pole dla ciekawości, natomiast eksperyment służy do ograniczania spekulacji do ram wyznaczanych przez rzeczywistość.

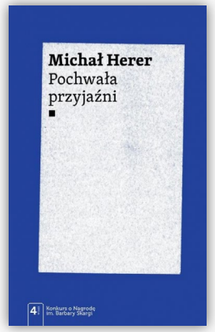
W trzeciej części książki Bartosz Brożek przygląda się pytaniu „dlaczego?" i omawia dwa podejścia do wyjaśniania w nauce: nomologiczno-dedukcyjne i mechaniczne. W tym pierwszym wyjaśnienie jakiegoś zjawiska polega na wskazaniu, że jest ono efektem jakiegoś ogólnego prawa, w drugim przypadku wyjaśnić coś to opisać mechanizm, który do tego czegoś doprowadził. Autor przestrzega przed zbyt pospiesznym i nieracjonalnym „zaspokajaniem" ciekawości, co może się przejawiać w dogmatyzmie. Esej kończy się psychologiczną analizą nudy.

Książka jest jedną z niewielu publikacji poświęconych wyłącznie zagadnieniu ciekawości. Szerokie spojrzenie autorów – od analiz filozoficznych i społecznych, przez przykłady z fizyki oraz omówienie badań psychologicznych – chociaż nie wyczerpuje tematu, to zapewnia doskonały punkt wyjścia do dalszych badań nad naturą tego zjawiska.

Piotr Biłgorajski

Bartosz Brożek, Michał Heller, Jerzy Stelmach, *Ciekawość*, Copernicus Center Press, Kraków 2023, 164 s.

Pochwała przyjaźni



Naturalnym sposobem istnienia człowieka jest bycie we wspólnocie. Życie razem to fundament społeczeństwa, pozwalający na rozwój i pogłębianie dobrobytu. Wraz z nastaniem nowoczesności i pojawieniem się jej rewolucyjnego postępu technologicznego tradycyjna, oparta na pokrewieństwie i wierzeniach religijnych wspólnota zaczęła przeżywać kryzys. Aby przetrwać, człowiek zaczął dbać przede wszystkim o własny szeroko rozumiany interes i, jak nazwał to zjawisko Michael Foucault, przybrał status „przedsiębiorcy samego siebie". Pojęcie rodziny jako triady ojciec – matka – dzieci oraz pojęcie państwa narodowego to wynalazki nowoczesności, które są lekarstwem na wywołaną przez ową nowoczesność alienację. Współcześnie rozumiana rodzina, której podstawą jest związek jednostek, pełni funkcję usprawiedliwienia udziału ludzi w konkurencyjnej grze wolnego rynku, gdzie fundamentalną zasadą jest rywalizacja z innymi o skończone dobra i pozycję. Ostatecznie bowiem walka z wrogim i obcym światem zewnętrznym stała się tym, co izoluje ludzi, zamykając ich w wąskiej wspólnocie.

Odpowiedzią na te wyzwania ma być proponowane przez Michała Herera, autora książki *Pochwała przyjaźni*, odnowienie *kultury przyjaźni*. Jego wyróżniony Nagrodą im. Barbary Skargi esej to inspirujący i pełen historiofilozoficznych odniesień, od Arystotelesa po Deleuze'a, namysł nad historią i obecnym stanem relacji przyjacielskiej oraz możliwym kierunkiem jej ekspansji. Przyjaźń rozumiana szeroko, jako siła przenikająca najróżniejsze relacje międzyludzkie, jest zdaniem autora niezwykle interesującym filozoficznie zjawiskiem, które nie dając się sprowadzić do innych typów relacji, takich jak np. miłość lub braterstwo polityczne, jednocześnie może w niesamowity sposób wpłynąć na owe relacje, nadając im nową jakość i sprawiając, że są dużo bardziej satysfakcjonujące i owocne. Przyjaźń jako przestrzeń do tworzenia wspólnoty, nieprowadząca jednak do homogeniczności, może nieść odpowiedź na niezwykle doniosłe pytanie: „Jak żyć razem?".

Paweł Sikora

Michał Herer, *Pochwała przyjaźni*, PWN, Warszawa 2017, 116 s.

Anegdoty i żarty

czyli filozofia na wesoło

W piątkowy wieczór 25 października 1946 roku Klub Nauk Moralnych Uniwersytetu Cambridge zorganizował spotkanie, którego głównym prelegentem był Karl Popper. Przybył on z Londynu, aby wygłosić odczyt zatytułowany *Czy istnieją problemy filozoficzne?* W gronie słuchaczy znaleźli się m.in. przewodniczący klubu, Ludwig Wittgenstein oraz Bertrand Russell. Popper stwierdził, że istnieją prawdziwe problemy filozoficzne, co wywołało sprzeciw Wittgensteina, uznającego je za jedynie językowe łamigłówki. W trakcie wymiany argumentów słynący z barwnej gestykulacji Wittgenstein używał pogrzebacza do wyakcentowania punktów swej wypowiedzi. Gdy Popper wymieniał istotne kwestie, Wittgenstein przerywał, twierdząc, że są to problemy logiki, nie filozofii. Popper w pewnym momencie odpowiedział, że istnieją także problemy etyczne. Wtedy Wittgenstein chwycił oparty o kominek pogrzebacz i wymachując nim wściekle, zwrócił się wzburzony do Poppera: „Daj nam przykład moralnego prawa!”. Niezrażony Popper odpowiedział z przekąsem: „Nie wygrażaj pogrzebaczem swoim gościom”. Po tych słowach Wittgenstein miał odrzucić pogrzebacz i opuścić salę, zatrzaskując drzwi z hukiem. Tak zdarzenie, które stało się legendą, relacjonował Popper. Niektórzy twierdzą nawet, że filozofowie pojedynkowali się na pogrzebacze.

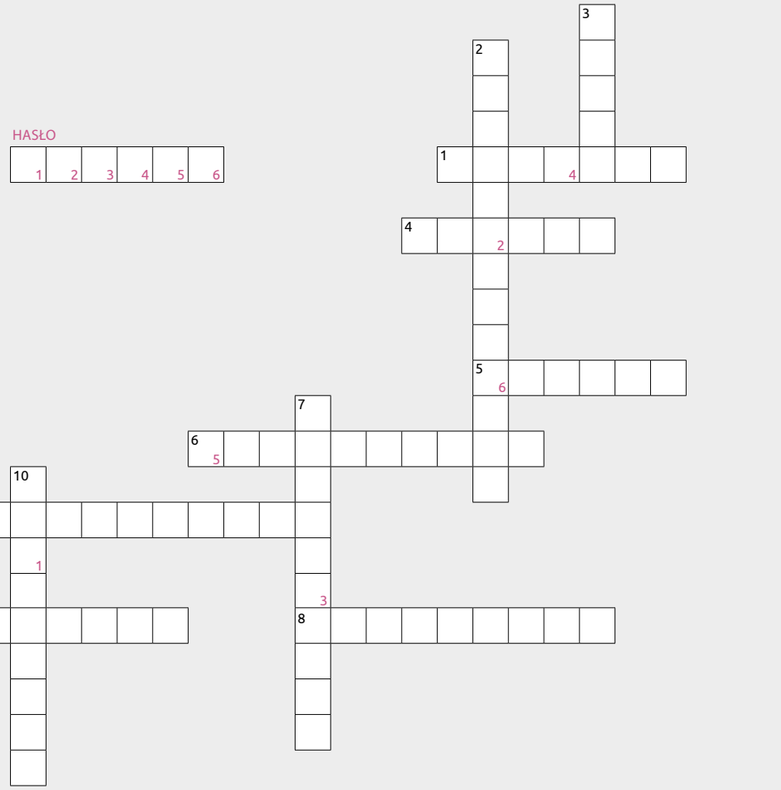
Źródło: D. Edmonds, J. Eidinow, *Pogrzebacz Wittgensteina*, Warszawa 2002.

Opracowanie: Milena Bartoszewska

Filozoficzna krzyżówka

czyli co zapamiętaliście z lektury tego numeru

1. Profesor fizyki i kosmolog, według którego do realizacji procesu poznawczego konieczna jest jedynie organizacja funkcjonalna materii, a nie jej określony rodzaj.
2. Jeden z problemów poruszanych w etyce AI, który dotyczy możliwości przewidywania i wyjaśniania algorytmów kierujących AI.
3. Określenie AI dysponującej umysłem podobnym do ludzkiego i wyznaczającej swoje własne cele.
4. Twórca teorii ewolucji biologicznej.
5. Badacz, według którego komputer jest kreatywny w stopniu porównywalnym do człowieka.
6. Jeden z problemów poruszanych w etyce AI, który dotyczy zbierania i analizowania danych o ludziach przez algorytmy.
7. W literaturze filozoficznej rozróżnia się ją fenomenalną i funkcjonalną.
8. Termin dotyczący „uczenia się” przez programy, uczenie...
9. Pojawienie się nowej własności bytowej na bazie występujących już elementów.
10. Nakaz rozumu w terminologii Immanuela Kanta.
11. Futurolog i transhumanista, który przestawił metaforę zatapiania ludzkich kompetencji przez ekspansję AI.



Opracowanie: Tomasz Kaliński

Rada naukowa: prof. dr hab. Krzysztof Brzechczyn; prof. dr hab. Adam Grobler; dr hab. Arkadiusz Gut, prof. UMK; prof. Jonathan Jacobs; prof. dr hab. Stanisław Judycki; prof. dr hab. Robert Piłat; prof. dr hab. Tadeusz Szubka; prof. Thomas Wartenberg; prof. dr hab. Jacek Wojtysiak

Kolegium redakcyjne: Piotr Biłgorajski, Elżbieta Drozdowska, Błażej Gębura, Robert Kryński, Marta Ratkiewicz-Siłuch (sekretarz redakcji), Artur Szutta (redaktor naczelny), Natasza Szutta, Mira Zyśko

Redaktor prowadzący numeru: Artur Szutta

Współpracownicy: Marcin Iwanicki, Aleksandra Pałka, Wojciech Rutkiewicz, Rafał Wąż

Oprac. redakcyjne i graficzne: Studio DTP Academicon | korekta: Piotr Biłgorajski, Elżbieta Drozdowska, Błażej Gębura, Dorota Krowicka, Marta Ratkiewicz-Siłuch, Agnieszka Stańczak, Mieszko Wandowicz; skład i grafika: ChatGPT, Patrycja Czerniak, Adam Dorot, Robert Kryński, Mira Zyśko; dtp@academicon.pl | dtp.academicon.pl

Dziękujemy naszym hojnym Patronom wspierającym nas na portalu Patronite.pl: Annie Bentyn, Filipowi Chmieleckiemu, Michałowi Markowi, Wojciechowi Malickiemu, Piotrowi Elfingerowi, Wojciechowi Grzegorczykowi, Katarzynie Frąckiewicz, Małgorzacie Laskowskiej, Sebastianowi Łasajowi, Janowi Swianiewiczowi, Zbigniewowi Szafrańcowi, Ani Wilk-Płaszczyk. Również dzięki Waszemu wkładowi możliwe było wydanie tego numeru.

Partnerzy medialni: Biznes na fali, Marka jest kobietą

Adres redakcji
Magazyn „Filozofuj!”
ul. H. Modrzejewskiej 13, 20-810 Lublin
e-mail: redakcja@filozofuj.eu
www: filozofuj.eu

Adres wydawcy
Wydawnictwo Academicon
ul. H. Modrzejewskiej 13, 20-810 Lublin
tel. 603 072 530, skype: academicon
e-mail: wydawnictwo@academicon.pl
www: wydawnictwo.academicon.pl

ISSN: 2392-2249
DOI: 10.52097/f.2024.1

Druk: Standruk

W następnym numerze…

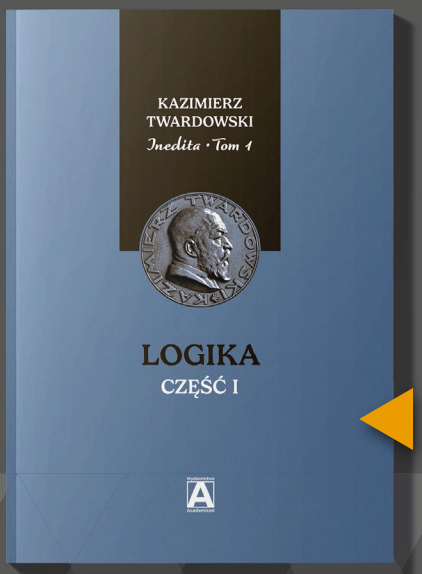
Dofinansowano ze środków Ministra Kultury i Dziedzictwa Narodowego pochodzących z Funduszu Promocji Kultury

Wydawnictwo Academicon

POLECANE KSIĄŻKI

Kazimierz Twardowski

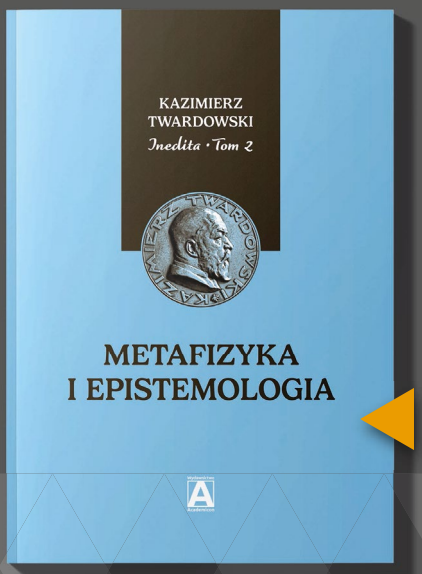
Inedita, t. 1:
Logika, cz. 1



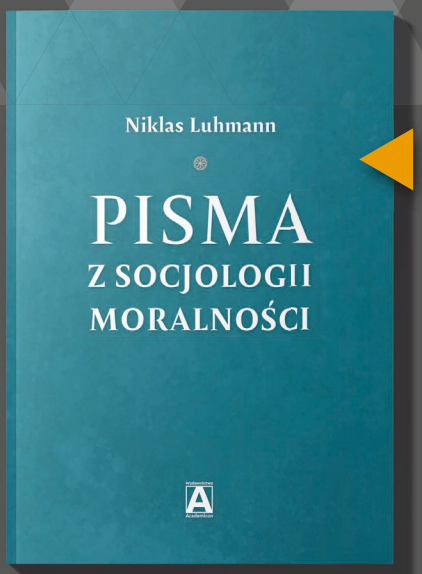
Cena: 63,00 zł
Oprawa twarda

Kazimierz Twardowski

Inedita, t. 2:
Metafizyka i epistemologia



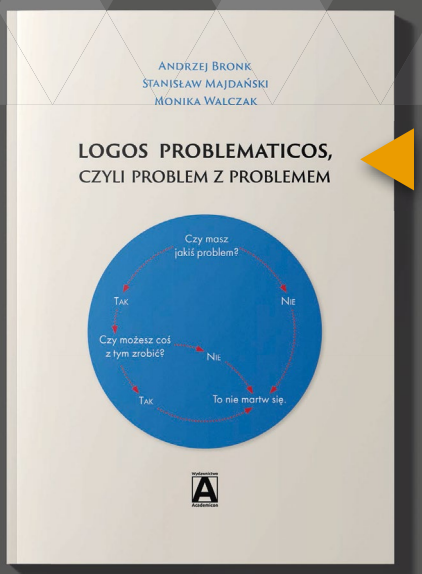
Cena: 63,00 zł
Oprawa twarda



Cena: 54,00 zł
Oprawa miękka

Niklas Luhmann

Pisma z socjologii moralności



Cena: 47,25 zł
Oprawa twarda

Andrzej Bronk, Stanisław Majdański, Monika Walczak

Logos problematicos, czyli problem z problemem

ZAKUPY: www: https://ksiegarnia.academicon.pl/, e-mail: ksiegarnia@academicon.pl, tel.: 603 072 530

Autopromocja