„Ethos" 36(2023) nr 3(143) 293-297

Robert KUBLIKOWSKI

FILOZOFIA SZTUCZNEJ INTELIGENCJI

Książka pod tytułem *Sztuczna inteligencja. Jej natura i przyszłość*[1] autorstwa Margaret A. Boden – opublikowana pierwotnie w Oxford University Press – dobrze wpisuje się w utrzymujące się od kilku dekad zainteresowanie tą problematyką[2].

Autorka monografii, o czym dowiadujemy się między innymi z informacji na okładce, jest profesorem nauk kognitywnych na Wydziale Informatyki Uniwersytetu w Sussex w Wielkiej Brytanii.

Polski, porządny przekład – wydany w serii wydawniczej „Krótkie Wprowadzenie" – został wykonany przez adiunkta Tomasza Sieczkowskiego, pracownika Wydziału Filozoficzno-Historycznego Uniwersytetu Łódzkiego. O naukową redakcję zadbał Piotr Fulmański, adiunkt na tamtejszym Wydziale Matematyki i Informatyki. Odredakcyjne przypisy objaśniają rzetelnie trudne fragmenty monografii.

W rozdziale pierwszym, „Czym jest sztuczna inteligencja?" (zob. s. 13-32), Boden wprowadza w tytułową problematykę. Zdaniem autorki sztuczna inteligencja (ang. artificial intelligence, AI), wykorzystuje różne techniki, których celem jest to, aby komputer wykonywał takie same czynności (rozumowanie i temu podobne), jak umysł ludzki. AI ma dwa cele główne: (1) technologiczny (praktyczny), kładący akcent na użyteczność komputerów i (2) naukowy (teoretyczny), polegający na zastosowaniu pojęć i modeli w taki sposób, żeby za pomocą AI można było odpowiedzieć na pytania dotyczące organizmów żywych (między innymi wykorzystywany przez biologów model „sztucznego życia", ang. A-life), a szczególnie odnoszące się do ludzi (zob. s. 13-15). To drugie zadanie jest mniej techniczne, a bardziej antropologiczne. Przy użyciu komputerów – ze względu na występowanie analogii między systemami informatycznymi a organizmami – można bowiem lepiej zrozumieć na przykład ludzkie umysły, ich naturę, strukturę, funkcje czy metody działania. Fundamentalnym pytaniem pozostaje, jakie są podobieństwa i różnice między inteligencją naturalną a sztuczną?

AI jest związana nie tyle z maszyną fizyczną (komputerem), ile z jakoś zdefiniowaną maszyną wirtualną, będącą symulacją analizowanego systemu. Maszyna wirtualna ma wykonywać procesy dokładnie w taki sposób, jak to się dokonuje w oryginalnym systemie fizycznym. Składa się ona ze wzorców działania – przetwarzania informacji (zob. s. 15-17).

Wyróżnia się pięć głównych typów AI: (1) klasyczna (symboliczna) AI (ang.

[1] Margaret A. B o d e n, *Sztuczna inteligencja. Jej natura i przyszłość*, tłum. T. Sieczkowski, red. P. Fulmański, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2020, ss. 208.

[2] Zob. Institute for Ethics in AI, www.oxford-aiethics.ox.ac.uk.

good old-fashioned AI, GOFAI). GOFAI zapoczątkował Alan Turing, wykazując, że wszystkie możliwe obliczenia mogą być zrealizowane przez system matematyczny, współcześnie określany jako uniwersalna maszyna Turinga. Przekonanie co do możliwości uzyskania AI wsparli neurolog i psychiatra Warren McCulloch oraz matematyk Walter Pitts, łącząc podejście Turinga z teorią synaps nerwowych autorstwa Charlesa Sherringtona i rachunkiem zdań Bertranda Russella. Arthur Samuel z kolei napisał program grający w szachy, który sam nauczył się wygrywać z jego twórcą. Natomiast Maszyna Teorii Logicznej (ang. logic theory machine), będąca w stanie udowadniać twierdzenia logiczne (Allen Newell, Herbert Simon), została prześcignięta przez Maszynę Rozstrzygającą Problemy (ang. general problem solver – Newell, John Shaw, Simon). Kolejnymi typami AI są: (2) sztuczne sieci neuronowe (koneksjonizm), (3) automaty komórkowe, (4) programowanie ewolucyjne i (5) systemy dynamiczne (zob. s. 18-32). Te typy AI autorka scharakteryzowała dokładniej w dalszej części swojej monografii.

Rozdział drugi, „Ogólna inteligencja jako Święty Graal" (zob. s. 33-69), dotyczy szansy utworzenia sztucznej inteligencji ogólnego przeznaczenia, czyli tak zwanej ogólnej sztucznej inteligencji (ang. artificial general intelligence, AGI). Gdyby udało się uzyskać AGI, to komputery byłyby w stanie funkcjonować nie tyle na mocy konkretnego, s z c z e g ó ł o w e g o oprogramowania dedykowanego do realizacji s z c z e g ó ł o w e g o celu, ile na podstawie możliwości percypowania, używania języka, rozumowania, kreatywności i tym podobnych w zastosowaniu do szerokiego, o g ó l n e g o spektrum problemów. Problemy mają być obliczalne efektywnie, czyli im mniej jest obliczeń, tym lepiej. Umożliwiają to następujące metody (strategie): (1) wyszukiwanie heurystyczne (akcent pada na takie zdefiniowanie programu, aby był on ukierunkowany na konkretne aspekty przestrzeni przeszukiwania), (2) planowanie (na odpowiednim poziomie abstrakcji może ono skutkować przycięciem – zredukowaniem drzewa reprezentującego elementy sprawdzone oraz te wymagające sprawdzenia), (3) upraszczanie matematyczne (między innymi milcząco uznawane jest założenie zaniedbujące rolę emocji w ludzkim działaniu) i (4) reprezentacja wiedzy w określonym języku programowania.

Reprezentacja wiedzy stosuje zbiór reguł JEŚLI-TO: JEŚLI dany Warunek jest spełniony, TO podejmij to Działanie.

Reprezentacja wiedzy może dotyczyć indywidualnych terminów (pojęć), a nie całych dziedzin wiedzy. Zdefiniowana hierarchiczna struktura (baza) danych to tak zwana r a m a. Problem ramy polega na tym, że programy AI nie mają ludzkiej zdolności do uchwycenia, rozumienia i s t o t n o ś c i, czyli tego, co jest ważne w konkretnej sytuacji. Problemu ramy można by uniknąć, gdyby były znane wszelkie możliwe k o n s e k w e n c j e każdego możliwego działania.

Reprezentacja wiedzy wymaga zdefiniowania stereotypowych sekwencji działań (na przykład określenie, co należy do „procedury" układania dziecka do snu).

W reprezentacji wiedzy akcentuje się tak zwane s i e c i s e m a n t y c z n e – terminy (pojęcia) pozostają w relacjach semantycznych, na przykład synonimiczności, antonimiczności, podporządkowania. Programy przeszukujące sieć internetową – silniki wyszukiwania – generalnie ujmując, programy NLP (ang. natural language processing) mogą znajdować związki między słowami czy tekstami, jednakże brak im r o z u m i e n i a tego, co przeszukują. Sieci semantyczne wiążą się również z poznaniem rozproszonym, czyli niezhierarchizowanym (holizm semantyczny).

Reprezentacja wiedzy – czy szerzej ujmując, AGI na poziomie człowieka – musi

uwzględniać także uczenie maszynowe (ang. machine learning). Wyróżnia się uczenie: (1) nadzorowane, (2) nienadzorowane i (3) przez wzmocnienie. W uczeniu nadzorowanym programista szkoli system przez zdefiniowanie – dla ustalonych danych wyjściowych – zbioru oczekiwanych rezultatów. System uczący się formułuje hipotezy dotyczące stosownych własności i odpowiednio – w zależności od szczegółowej informacji zwrotnej o błędach – koryguje hipotezę wyjściową. W uczeniu nadzorowanym użytkownik nie określa oczekiwanych rezultatów i nie otrzymuje komunikatów o błędach. Przyjmowana jest zaś reguła, że własności współwystępujące będą współwystępowały również w przyszłości. Natomiast uczenie przez wzmocnienie wykorzystuje schemat nagród (za sukces) i kar (za porażkę): system – na podstawie komunikatów zwrotnych – jest informowany, że jego dana reakcja była poprawna lub niepoprawna. Wyniki są aktualizowane po pojedynczej decyzji lub sekwencji decyzji.

W rozdziale trzecim, „Język, kreatywność, emocje" (zob. s. 71-91), została zaakcentowana rola NLP w tłumaczeniu maszynowym i trudności związane z analizami, zarówno syntaktyczną, jak i semantyczną. Niuanse struktury wypowiedzi powodują zmianę sensu i referencji. Oprócz tego w użyciu języka ważny jest również aspekt pragmatyczny: adekwatność (relewancja) użycia wyrażenia w danej sytuacji i kontekst użycia. Kontekst jest tworzony między innymi przez (1) wyrazy bliskoznaczne, (2) synonimy, (3) antonimy, (4) wyrazy, których zbiory (zakresy) są w relacji zawierania się w jakieś klasie, a także (5) wyrazy, których desygnaty są w relacji należenia do jakiejś klasy. Kontekstualne są dla siebie również (6) wyrazy, których desygnaty są w relacji część–całość. Częstość współwystępowania słów jest ważna przy wyszukiwaniu (eksplorowaniu) informacji.

Pojęcia związane z AI są użyteczne w wyjaśnianiu i porządkowaniu (typologizowaniu) ludzkiej kreatywności. Wyróżnia się: (1) kreatywność kombinacyjną – polegającą na tym, że znane idee są łączone w dotychczas nieznane sposoby, (2) kreatywność eksploracyjną – bazującą na kulturowo znanych sposobach (stylach, regułach) myślenia i ich niedużych zmianach, (3) kreatywność transformacyjną – dopuszczającą duże zmiany reguł w celu uzyskania radykalnej nowości.

Do dowartościowania emocji w AI przyczynili się Marvin Minsky i Aaron Sloman, którzy postulowali, aby traktować umysł jako całość (intelekt-racjonalność, wola-decyzja, emocje).

Rozdział czwarty, zatytułowany „Sztuczne sieci neuronowe" (zob. s. 93-116), scharakteryzował je jako układy złożone z licznych, wzajemnie połączonych jednostek (elementów) przetwarzających. Każda z nich wykonuje tylko jeden rodzaj obliczeń. Jedną ze sztucznych sieci neuronowych jest przetwarzanie równoległe rozproszone (ang. parallel distributed processing, PDP). Zaletą PDP jest: (1) umiejętność uczenia się wzorców i związków między wzorcami za pośrednictwem przykładów, a nie przez bezpośrednie programowanie, (2) tolerancyjność względem nieprecyzyjnych informacji, (3) możliwość rozpoznawania niekompletnych lub częściowo uszkodzonych wzorców, (4) odporność na zniszczenia. Jeżeli uszkodzenia są coraz większe, to działanie PDP pogarsza się stopniowo (zob. s. 93-101).

System – w tak zwanym uczeniu głębokim (ang. deep learning) – przyswaja sobie strukturę głęboką, a nie tylko wzorce powierzchowne występujące w jakiejś dziedzinie (zob. s. 102-116). Ta problematyka łączy się z – dyskutowanymi w filozofii języka i nauki – esencjalizmem, teorią rodzajów natu-

ralnych wraz z teorią terminów naturalno-rodzajowych, a także z teorią definicji ostensywnych i realnych (istotowych) oraz problemem tak zwanej r a m y.

W rozdziale piątym, ,,Roboty i sztuczne życie" (zob. s. 117-134), autorka podkreśla, że roboty k l a s y c z n e z AI potrafiły naśladować ś w i a d o m e d z i a ł a n i a człowieka. Nie były jednak w stanie reagować natychmiast na zmiany w otoczeniu. Natomiast roboty współczesne – tak zwane roboty u s y t u o w a n e – są oparte na r e a k c j i n a s y t u a c j ę, na zachowaniu adaptacyjnym. Zdolność do odruchów sensomotorycznych jest zasadniczo zapewniona dzięki anatomii robota (jego czujnikom i połączeniom sensomotorycznym), a nie jego oprogramowaniu.

Roboty – zwane też automatami komórkowymi – są systemami poszczególnych jednostek, z których każda przybiera jeden ze skończonej ilości stanów odpowiednio do prostych reguł zależnych od bieżącego stanu sąsiadujących jednostek składających się na automat (twórcą automatów komórkowych był John von Neumann).

Roboty, w których implementowane jest programowanie ewolucyjne (możemy to prześledzić w pracach na przykład Norberta Wienera), mogą się zmieniać (rewidować i korygować) dzięki algorytmom genetycznym inspirowanym przez biologię ewolucyjną.

Rozdział szósty, pod tytułem „Ale czy to naprawdę jest inteligencja?" (zob. s. 135--160), zawiera fundamentalne pytania na temat AI: Czy przyszłe systemy AGI będą dysponowały prawdziwą świadomością, inteligencją, rozumieniem i kreatywnością? Czy będą one posiadały status moralny i możliwość wolnego wyboru? Próbowano udzielić na nie odpowiedzi (na przykład test Turinga czy myślowy eksperyment Johna Searle'a o tak zwanym chińskim pokoju obrosły dyskusjami), lecz nie dostarczyły one bezspornych stanowisk.

W ostatnim rozdziale (siódmym), „Osobliwość" (zob. s. 161-183), zasygnalizowano dylematy moralne związane ze sztuczną inteligencją. Rozwój technologiczny AI zastępującej ludzi może zwiększyć bezrobocie, ale też skutkować pojawieniem się nowych zawodów i miejsc pracy. Wątpliwości etyczne budzi również konstruowanie robotów erotycznych (sztucznych partnerów seksualnych) czy robotów do opieki nad ludźmi niepełnosprawnymi. Czy sztuczna empatia jest moralnie dobra i akceptowalna? Czy budowanie i używanie takich robotów jest etycznie uzasadnione? Podobny dylemat dotyczy robotów militarnych, skonstruowanych do likwidacji celów wskazanych przez obsługujący je personel, a nawet wybranych samodzielnie przez robota. Poszukiwane są kodeksy postępowania dla robotów, mające rozwinąć „kompetencje moralne" AI. Przykładem są „trzy prawa robotyki" Issaca Asimova: (1) robot nie powinien skrzywdzić człowieka, (2) robot musi wypełniać ludzkie polecenia, (3) robot powinien dbać o własne przetrwanie. W przypadku konfliktu interesów pierwszeństwo ma prawo pierwsze. Ideałem byłoby utworzenie systemu obliczeniowego zdolnego do przeprowadzania rozumowań i dyskusji moralnych.

Książka została zaopatrzona także w obszerną „Bibliografię" (zob. s. 185-199) i rozbudowany „Indeks osobowo-rzeczowy" (zob. s. 201-208).

Podsumowując, styl pisarski autorki monografii mógłby być niekiedy precyzyjniejszy (na przykład w początkowym rozdziale badaczka nie wskazuje dokładnie, które typy AI są reprezentowane przez których twórców). Jednak nie jest to częste. Recenzowana książka zasadniczo jest wartościowym – merytorycznym, erudycyjnym, a jednocześnie wyważonym i zwartym – wprowadzeniem w zagadnienia związane z jednej strony z informatyką, a z drugiej − z filozofią nauki i techniki, a dokładniej z filozofią (ontologią,

semiotyką, metodologią, epistemologią czy etyką) sztucznej inteligencji. Dyscypliny te mogą być dla siebie nawzajem inspiracją do badań uwzględniających nowe konteksty i wyzwania teoretyczne oraz praktyczne.

Kontakt: Katedra Metodologii Nauk, Instytut Filozofii, Wydział Filozofii, Katolicki Uniwersytet Lubelski Jana Pawła II, Al. Racławickie 14, 20-950 Lublin
E-mail: robert.kublikowski@kul.pl
ORCID: 0000-0002-5199-3941