Gniewosz Leliwa

INTELIGENTNY INACZEJ

Czy AI naprawdę myśli,
czy tylko doskonale udaje?

Helion

Redaktor prowadzący: Małgorzata Kulik

Projekt okładki: Studio Gravite/Olsztyn
Obarek, Pokoński, Pazdrijowski, Zaprucki
Grafika na okładce została wykorzystana za zgodą AdobeStock.com.

Helion S.A.
ul. Kościuszki 1c, 44-100 Gliwice
tel. 32 230 98 63
e-mail: *helion@helion.pl*
WWW: *helion.pl* (księgarnia internetowa, katalog książek)

Drogi Czytelniku!
Jeżeli chcesz ocenić tę książkę, zajrzyj pod adres
helion.pl/user/opinie/ininai
Możesz tam wpisać swoje uwagi, spostrzeżenia, recenzję.

ISBN: 978-83-289-4195-3

Printed in Poland.

SPIS TREŚCI

SŁOWEM WSTĘPU

Problem ze wstępem pisanym na samym końcu jest taki, że wszystko, co miałem do napisania, już napisałem… i teraz muszę rzeźbić. Podobno są ludzie, którzy od wstępu zaczynają, ale ja, szczerze mówiąc, nigdy nikogo takiego nie spotkałem. Podejrzewam, że to mogą być ci sami ludzie, którzy wierzą, że dyskusje w internecie są po to, żeby kogokolwiek do czegokolwiek przekonać. Urocze. No cóż, takie podejście mimo całego szeregu oczywistych wad ma też jedną niezaprzeczalną zaletę: będzie krótko i zwięźle. W kilku słowach wyjaśnię, po co i dla kogo jest ta książka, a potem już bez zbędnego przedłużania zapraszam na samo słodkie.

Po pierwsze trzymacie w rękach esej popularnonaukowy, co oznacza, że oprócz rzetelnej wiedzy podanej w przystępny sposób — bez udawanej skromności — znajdziecie tu również wiele moich osobistych przemyśleń i refleksji. Za cel postawiłem sobie znalezienie odpowiedzi na nurtujące mnie od dłuższego czasu pytanie: czy sztuczna inteligencja naprawdę jest inteligenta i czy w ogóle da się odtworzyć ludzką inteligencję w maszynie?

Odpowiedź wydaje się oczywista, prawda? Każdy, kto miał przyjemność pracować z AI, wie, jak imponujące rzeczy potrafi robić. A skoro w popularnych benchmarkach radzi sobie lepiej niż większość z nas, lepiej zdaje egzaminy i olimpiady przedmiotowe, to znaczy, że już nie tylko nam dorównała, ale w wielu dziedzinach wręcz nas prześcignęła. Jeśli tak pomyśleliście, to strasznie się cieszę, że ta książka trafiła właśnie w Wasze ręce. Nie zrozumcie mnie źle, uwielbiam przekonywać przekonanych, ale to za tych nieprzekonanych są prawdziwe punkty.

Ale, ale… przekonani też znajdą tu coś dla siebie. Postanowiłem bowiem schować ego do kieszeni, szczerze przyznać przed samym sobą, jak niewiele wiem, i z dziecięcą radością odkrywać świat na nowo. Nie tylko dla Was, ale też razem z Wami. Mój romans ze sztuczną inteligencją trwa już tyle lat, że za moment będzie mógł legalnie napić się piwa, co nie zmienia faktu, że o psychologii, neurobiologii czy filozofii wiedziałem do tej pory znacznie mniej, niż chciałbym przyznać. Nie twierdzę, że zostałem ekspertem w każdej z tych dziedzin, ale pracę domową uczciwie odrobiłem.

Po drugie postawiłem na absolutną uczciwość wobec Was i wobec siebie. Nie ukrywam, że do pisania siadałem z całym bagażem własnych doświadczeń, poglądów i przekonań, których nie sposób całkowicie się wyzbyć. I choć w mojej dotychczasowej działalności popularyzatorskiej dałem się poznać raczej jako sceptyk-realista niż entuzjasta, to nie pisałem pod tezę — to mogę Wam obiecać. Potraktowałem pisanie jak podróż w nieznane, na którą wyruszam z otwartym umysłem, a nie jak okazję do udowodnienia komukolwiek swoich racji. Myślę, że w głębi duszy chciałem się mylić i… w wielu przypadkach rzeczywiście tak było. Oczywiście nie zdradzę Wam zakończenia, ale jestem pewien, że nieraz uda mi się Was zaskoczyć, zadziwić i… rozbawić.

Konieczność zgłębienia podstaw filozofii, psychologii i neurobiologii w kontekście sztucznej inteligencji była dla mnie doskonałą okazją do ciągłego pytania: dlaczego? Pisałem, czytałem to, co napisałem, a potem pytałem, czy to, co napisałem, naprawdę mnie przekonuje. Jeśli nie, to wałkowałem temat z każdej możliwej strony tak długo, aż w końcu z satysfakcją stwierdzałem, że został wyczerpany. Albo zaczynałem od zera. Bez sentymentów i bez skrupułów.

Po trzecie i najważniejsze: myślę, że jako dorośli zbyt rzadko pozwalamy sobie na dziecięce zdziwienie, zachwyt i chęć odkrywania, dlaczego świat działa akurat tak, a nie inaczej. W natłoku codziennych obowiązków zamiast pytać i dociekać, wolimy wzruszyć ramionami i uznać, że tak po prostu jest. Ja, pisząc tę książkę, postanowiłem być znowu dzieckiem, dla którego „bo tak" nie jest żadną odpowiedzią. Wy, czytając ją, odkryjcie w sobie na nowo tę cudowną, dziecięcą ciekawość i nigdy, przenigdy nie pozwólcie, by proza codzienności Wam ją odebrała…

MYŚLĘ, WIĘC KOMPLIKUJĘ

Czymże więc jest inteligencja? Jeśli nikt mnie o to nie pyta, wiem. Jeśli pytającemu usiłuję wytłumaczyć, nie wiem.

Myślę, że ta parafraza słów św. Augustyna z Hippony, w oryginale odnoszących się do czasu, świetnie oddaje fakt, że intuicyjnie wszyscy pojęcie inteligencji jakoś rozumiemy, a mimo to mamy ogromne trudności ze spójnym jej zdefiniowaniem.

Nie mogę rozpocząć tej książki inaczej niż od próby uporządkowania wiedzy na jej temat. Wiem, wiem, to strasznie niepopularne podejście, żeby najpierw uzgadniać pojęcia, a dopiero potem o nich dyskutować. Zwłaszcza że współczesny świat z platformami społecznościowymi na czele uwielbia clickbait, spór, oburzenie i prosty, jednoznaczny przekaz. Trudno, to moja książka, więc będę ją pisał tak, jak uznam za stosowne.

Nie sposób rozmawiać o sztucznej inteligencji, nie wiedząc, czym jest ta prawdziwa. Wydaje się bowiem, że obie inteligencje powinny być w zasadzie tożsame, a przymiotnik „sztuczna", zgodnie z definicją *Słownika języka polskiego PWN*, powinien odnosić się do bycia stworzonym przez człowieka w celu zastąpienia naturalnego odpowiednika.

Problem w tym, że jednej, spójnej i powszechnie akceptowanej definicji inteligencji po prostu nie ma. Istnieje wiele różnych, lepszych i gorszych, a tej jednej jedynej brak. Być może do tej pory wcale jej nie potrzebowaliśmy, bo i żadnej inteligencji poza ludzką po prostu nie znaliśmy?

Zamiast wykładu *ex cathedra* proponuję dużo swobodniejszą konwencję — zabawimy się w detektywa, który próbuje ustalić fakty na podstawie dostępnych

dowodów i materiałów ze śledztwa. Dedukcja, podejrzliwość i myślenie krytyczne — oto nasze narzędzia. Zabawa może być pouczająca szczególnie dla młodszego pokolenia, dla którego domyślnym odruchem staje się wklepanie hasła w ChatGPT i przyjęcie wygenerowanych treści na wiarę. O nie, tutaj tak nie robimy.

***

Zacznijmy od czegoś prostego. Wspomniany wcześniej *Słownik języka polskiego PWN* definiuje inteligencję jako zdolność rozumienia, uczenia się oraz wykorzystywania posiadanej wiedzy i umiejętności w sytuacjach nowych. Wszystko fajnie, ale jak rozumieć rozumienie? Ten sam słownik podpowiada: rozumieć znaczy pojmować treść przekazywanej informacji, zdawać sobie sprawę z czegoś lub przypisywać czemuś określone znaczenie.

I tu zaczynają się schody. Czy musimy rozumieć polecenie, by móc je wykonać? Czy komputer rozumie treść algorytmu, czy tylko przetwarza kod do coraz prostszej postaci, aż do napięcia na elektrodach w tranzystorze? Czujemy chyba, że ten aspekt to ślepa uliczka. Jeśli raz stwierdzimy, że maszyna nie rozumie sortowania bąbelkowego (jeden z najprostszych algorytmów porządkujących dane), to musimy być konsekwentni, uznając, że to samo dotyczy wykonywania znacznie bardziej skomplikowanych programów, wliczając w to duże modele językowe.

Zdaję sobie również sprawę, że w tym miejscu muszą pojawić się argumenty w stylu: nasz mózg to też komputer, a całe to myślenie i rozumienie sprowadza się do sygnałów elektrochemicznych w neuronach. Czym to niby różni się od napięcia na elektrodach w tranzystorze? Obiecuję, że jeszcze wrócimy do tych zagadnień, ale na razie zaparkujmy je w tym miejscu, bo nie wydaje się, żeby rozumienie przybliżało nas w jakiś satysfakcjonujący sposób do naszej superdefinicji inteligencji.

Może w takim razie zdolność uczenia się? To możemy zdefiniować jako proces poznawczy prowadzący do modyfikacji zachowania osobnika pod wpływem doświadczeń, co zwykle zwiększa przystosowanie osobnika do otoczenia. Polska Wikipedia podpowiada nam, że zdolność uczenia się w różnym zakresie posiadają zwierzęta, ludzie, grupy ludzi, a także… komputery.

Komputery? A jakże! Istnieje cały ogromny dział sztucznej inteligencji poświęcony algorytmom, które automatycznie poprawiają swoje działanie poprzez doświadczenie, czyli poprzez ekspozycję na dane. To uczenie maszynowe, którego początków należy doszukiwać się w latach 50. XX wieku.

Weźmy na tapet systemy rekomendacyjne, choćby algorytm Netfliksa, które w oparciu o nasze wcześniejsze wybory uczą się, jak sugerować nam coraz to bardziej dopasowane do nas produkty i usługi (np. filmy i seriale). Nie mamy chyba wątpliwości, że te systemy się uczą, prawda? Zgodnie z definicją poprawiają swoje zachowanie pod wpływem doświadczeń, takich jak obejrzane przez nas filmy, ich oceny oraz czas spędzany na oglądaniu konkretnych produkcji. Czy to jednak wystarczy, żebyśmy nazwali je inteligentnymi?

Możemy się chyba zgodzić, że to ciut za mało. Tak jak uczenie maszynowe jest częścią sztucznej inteligencji, tak zdolność uczenia się jest istotnym, ale nie jedynym elementem inteligencji. Nie wystarczy, by zdefiniować ją w sposób jednoznaczny i wyczerpujący.

Warto przy tej okazji pochylić się nad przystosowywaniem się osobnika do otoczenia, czyli nad jego zdolnością do adaptacji. To element, który przewija się zarówno przy próbie zdefiniowania inteligencji, jak i uczenia się. Element o tyle problematyczny, że nie wyklucza choćby takiego termostatu, który poprzez aktywne działanie utrzymuje w pomieszczeniu zadaną temperaturę.

O nie, teraz to przesadziłeś. Przecież wiadomo, że termostat to za mało. W porządku, a co z organizmami jednokomórkowymi? Czy ruch, który w odpowiedzi na bodźce pozwala im oddalić się od szkodliwych substancji (np. chemikaliów), wystarczy? Gdzie postawimy granicę? Bo jeśli jej nie postawimy, to z przykrością informuję, że szlag trafi naszą superdefinicję.

Wygląda na to, że nie znajdziemy pojedynczej cechy, która w sposób jednoznaczny świadczyłaby o inteligencji jej posiadacza. Być może kluczem do sukcesu jest połączenie kilku różnych cech, na przykład pozyskiwania wiedzy i umiejętności w procesie uczenia się oraz wykorzystywania nowo pozyskanej wiedzy i umiejętności w zupełnie nowych warunkach? Gdyby algorytm Netfliksa, stwierdzając, że nie znamy języka hiszpańskiego, nagle zaczął nas go uczyć, żebyśmy mogli obejrzeć sobie *Dom z papieru* w oryginale, to byłoby coś, prawda?

Zabawmy się w adwokata diabła. Czy gdy poprosimy chatbota o napisanie wiersza o nas, podając mu szczegóły z naszego życia, do których nie miał wcześniej dostępu, to czy należałoby to uznać za pozyskanie nowej wiedzy i zastosowanie jej w nowej sytuacji?

Z pewnością znajdą się osoby, które powiedzą, że tak. I trudno im się dziwić. Ale czy napisanie wiersza nie jest specyficznym przypadkiem zastosowania wiedzy i umiejętności, które chatbot posiadał od samego początku? Został w końcu wytrenowany do generowania treści w oparciu o zadaną instrukcję (prompt). Nie nauczył się czegoś, czego nie umiał już wcześniej. Informacje na mój temat nie zmieniły parametrów jego ogromnej sieci neuronowej. Owszem, może je przechowywać do końca świata i wykorzystać w razie potrzeby jako kontekst dla kolejnych promptów, ale to tylko prosta sztuczka, którą ciężko nazwać pozyskaniem nowej wiedzy.

Wrócimy do tych zagadnień, obiecuję. Na razie zależało mi jedynie na tym, żeby pokazać, że powinniśmy być bardzo ostrożni, gdy mówimy o wiedzy i umiejętnościach w kontekście dużych modeli językowych. Umiejętność posługiwania się językiem naturalnym jest dla nas tak silnie skorelowana z inteligencją, że niemal odruchowo traktujemy ją jako warunek wystarczający do bycia inteligentnym. Takie podejście może nas bardzo łatwo zwieść na manowce.

***

Zerknijmy jeszcze do *Encyklopedii PWN*. Przeczytamy w niej, że inteligencja jest jednym z najbardziej wieloznacznych pojęć w psychologii, odnoszącym się do sprawności w zakresie czynności poznawczych. Nie pomagasz, encyklopedio, oj, nie pomagasz.

Dalej, oprócz tego, co już wiemy, natrafiamy na coś bardzo interesującego: inteligencja ujmowana jako cecha ludzkiego umysłu to zdolność myślenia, rozwiązywania problemów oraz angażowania adekwatnych do okoliczności procesów poznawczych, od których zależy skuteczność przystosowania się do nowych sytuacji i sprawność działania. Inteligencja jako cecha ludzkiego umysłu. Ciekawe, prawda?

Zostawmy na razie myślenie, bo zapewne zaprowadzi nas tam, gdzie wcześniej rozumienie. Rozwiązywanie problemów nie wydaje się szczególnie interesujące

z naszej perspektywy. W zasadzie każde zastosowanie sztucznej inteligencji służy rozwiązaniu jakiegoś bardzo konkretnego problemu, np. rekomendacji filmu czy znalezieniu najkrótszej drogi z punktu A do punktu B. Ma to zresztą swoją nazwę: słaba (lub wąska) sztuczna inteligencja to taka, która rozwiązuje tylko te problemy, do których została przyuczona. I żadne inne. I wbrew przeróżnym opiniom tylko taką sztuczną inteligencję umiemy wytwarzać.

A co z angażowaniem adekwatnych do okoliczności procesów poznawczych? Brzmi groźnie, ale tak naprawdę chodzi o to, żeby nie strzelać do muchy z armaty. To dość istotny aspekt, zwłaszcza w kontekście ekspozycji na bodźce wymaganej do przyswojenia sobie nowej wiedzy lub umiejętności. Dziecko, gdy raz zobaczy kotka, świetnie poradzi sobie z rozpoznawaniem tych zwierząt w przyszłości. Sztuczne sieci neuronowe potrzebują zobaczyć setki czy tysiące różnych kotów, żeby opanować tę samą umiejętność.

Sztuczna inteligencja jest bardzo nieefektywna w wykorzystywaniu dostępnych zasobów w procesie uczenia się. Jeśli trzymać się porównania z muchą, to w jej przypadku do eksterminacji owada wysyłamy nie armatę, lecz całe wojsko. Czy nazwiemy inteligentnym kogoś, kto wkuwa na blachę? Nasze szkolne doświadczenie podpowiada, że nie.

Testy mierzące iloraz inteligencji (IQ), abstrahując od ich uzasadnionej krytyki, często sprowadzają się do rozwiązywania zadań, w których badanemu przedstawia się kilka pierwszych elementów jakieś sekwencji, a jego zadaniem jest odkrycie wzorca i ustalenie, jaki powinien być jej kolejny element. Znalezienie rozwiązania możemy traktować jak nabycie nowej umiejętności przy bardzo ograniczonej ekspozycji na bodźce.

Do tego też jeszcze wrócimy, a na razie odnotujmy, że choć raczej nie jest to cecha, która rozstrzygałaby o inteligencji lub jej braku, to może być całkiem użyteczna w różnego rodzaju porównaniach. Mozolna i nieefektywna nauka to wciąż nauka, ale czy ktoś, kto w okamgnieniu odkrywa złożone reguły rządzące niewielkim zbiorem danych, jest tak samo inteligentny jak ktoś, kto musi zobaczyć tysiące podobnych zadań, żeby dojść do analogicznych wniosków?

***

Wygląda na to, że zerwaliśmy już wszystkie nisko wiszące owoce, a nasze śledztwo nie przybliżyło nas specjalnie do jednej upragnionej superdefinicji. Chyba pora na wytoczenie cięższych dział: psychologów, którzy zjedli na tym zęby, i teorii, które chcieliby nam zaproponować.

Zacznijmy od Charlesa Spearmana, który twierdził, że wszystkie zdolności człowieka składają się z dwóch czynników: inteligencji ogólnej (czynnik g) oraz zdolności specyficznych (czynnik s). Od razu zaznaczę, że nie będziemy zajmować się zdolnościami specyficznymi. Są niezależne od inteligencji ogólnej i obejmują twarde kompetencje (np. umiejętność gry w szachy i czytanie nut) potrzebne do wykonania konkretnych zadań. Nie przydadzą się w naszym śledztwie.

Nas interesuje inteligencja ogólna, nazywana również energią umysłową, która leży u podstaw wszystkich działań poznawczych. Mamy tu nabywanie i rozumienie własnego doświadczenia, obejmujące umiejętności obserwowania tego, co dzieje się w naszym umyśle, oraz myślenie abstrakcyjne, obejmujące umiejętności wydobywania związków lub konsekwencji z kilku zdarzeń, bodźców lub przesłanek.

Przekładając to na język konkretu, można stwierdzić, że Spearman określa inteligencję jako zdolność dostrzegania zależności między różnymi pojęciami oraz wyciągania wniosków na podstawie tych zależności. Ta pierwsza, zwana również edukcją relacji, to zrozumienie, że w związku „pies — szczekanie" chodzi o zwierzę i dźwięk, który ono wydaje. Ta druga, zwana edukcją korelatu, umożliwia ustalenie, że w analogicznej relacji „kot — ?" brakującym elementem jest „miauczenie". Aha, edukcja, nie edukacja — tam nie ma „a", którą uporczywie wstawia nasz mózg!

Brzmi to zaskakująco podobnie do odkrywania wzorców w zadanej sekwencji, prawda? Wciąż krążymy wokół tego samego, ale teraz przynajmniej używamy mądrych słów i mieliśmy przyjemność poznać pana Spearmana. Dla porządku dodam, że twórca dwuczynnikowej teorii inteligencji uznawał czynnik ogólny za zdeterminowany genetycznie — niezmienny w ciągu życia i inny dla każdej jednostki.

Robert Sternberg zaproponował triarchiczną (tak, wiem) teorię inteligencji, która dzieli ją na trzy rodzaje: analityczną, twórczą i praktyczną. Jego zdaniem inteligencja jest relacją jednostki z kontekstem (światem zewnętrznym lub wewnętrznym) na poziomie poznawczym. Zdolnością umysłową do celowej adaptacji, selekcji

lub zmiany w otoczeniu. To ostatnie jest szczególnie interesujące, gdyż rozwiązuje nam problem termostatów i organizmów jednokomórkowych reagujących na bodźce.

Inteligencja analityczna to myślenie krytyczne i rozwiązywanie problemów za pomocą logiki. Wyciąganie wniosków, używanie reguł i wzorów. Umiejętność rozkładania złożonych problemów na czynniki pierwsze. Planowanie, ocenianie i porównywanie.

Inteligencja twórcza to myślenie nieszablonowe i radzenie sobie w nowych sytuacjach. Umiejętność generowania nowych pomysłów, rozwiązań i spojrzeń na problemy. Kreatywne łączenie informacji z różnych dziedzin. Wychodzenie poza to, co znane i oczywiste.

Inteligencja praktyczna to zdrowy rozsądek (albo chłopski rozum). Wykorzystywanie wiedzy i doświadczenia w codziennym życiu. Dopasowanie swojego zachowania do sytuacji lub wręcz przeciwnie — kształtowanie środowiska w taki sposób, aby sprzyjało naszym celom.

Spojrzenie Sternberga jest o tyle cenne, że zdaje się wykraczać poza nasze dotychczasowe rozważania. Inteligencja to nie tylko część analityczna, ale też twórcza i praktyczna. Obawiam się jednak, że zaproponowany trójpodział przyniesie nam więcej pytań niż odpowiedzi. Czy sztuczna inteligencja jest twórcza, gdy tworzy nowe kombinacje znanych elementów? Czy rekombinacja istniejących wzorców jest tym samym co ludzka kreatywność? Czy przy braku zmysłów, bezpośredniego doświadczenia i możliwości wpływania na świat realny w ogóle możemy mówić o inteligencji praktycznej? Odnotujmy te wszystkie pytania i przejdźmy do kolejnego wielkiego badacza.

A jest nim Howard Gardner, który zasłynął swoją teorią inteligencji wielorakiej. Tym razem dzielimy ją nie na dwa czy trzy, ale aż na osiem odrębnych bloków. Zdaniem Gardnera możemy wyróżnić następujące typy inteligencji: logiczno-matematyczną, językową, przyrodniczą, muzyczną, przestrzenną, cielesno-kinestetyczną, interpersonalną oraz intrapersonalną.

Z każdą z nich możemy skojarzyć konkretny kierunek rozwoju, np. przyrodnicza to biolog, muzyczna — kompozytor, a cielesno-kinestetyczna — sportowiec, a także

kluczowe komponenty, np. inteligencja interpersonalna odpowiada za umiejętność dostrzegania i właściwego reagowania na nastroje, temperament, motywacje i pragnienia innych ludzi.

Chyba każdy bez większych trudności wymieni kierunki i kluczowe komponenty skorelowane z poszczególnymi typami inteligencji wielorakiej. Czy taki podział ma sens? Trudno powiedzieć. Z pewnością nie cieszy się on powszechnym uznaniem świata nauki i — z bardzo różnych powodów — pozostaje obiektem krytyki wielu psychologów.

Z jednej strony wskazuje na nowe, interesujące obszary, o których nie pomyśleliśmy wcześniej, jak choćby inteligencja przestrzenna. Niektórzy z nas są świetni w poruszaniu się po mieście bez mapy. Robią to instynktownie i rzadko się mylą. Niektórzy z nas jeszcze w szkole na lekcjach geometrii bez pudła potrafili wskazać, jaki kształt będzie miał przekrój dowolnej bryły dowolną płaszczyzną. Nie zapamiętywali przecież wszystkich możliwych konfiguracji, ale tworzyli w swojej głowie efektywny model, na którym swobodnie operowali w myślach. Zapamiętajmy te obserwacje, gdyż przydadzą się nam w nieodległej przyszłości.

Z drugiej jednak strony podział na wiele niezależnych bloków zdaje się sugerować, że wystarczy wykazać się jednym, dowolnym typem inteligencji, by zasłużyć na miano inteligentnego. Czy roboty z Boston Dynamics są inteligentne, ponieważ świetnie radzą sobie z typem cielesno-kinestetycznym? Czy ChatGPT jest inteligentny, bo opanował typ językowy lepiej niż niejeden człowiek? I czy rzeczywiście sposób, w jaki generuje treści, wystarczy, by nazwać go inteligentnym językowo?

Czy to już wszyscy? A gdzie tam! Obawiam się, że nie jesteśmy nawet w połowie stawki, ale nie będziemy omawiać wszystkich teorii i ich twórców. Dla porządku wymienię jeszcze kilka nazwisk, które równie dobrze mogłyby paść w tym rozdziale: William Stern, Louis Thurstone, David Wechsler, Jean Piaget, Joy Guilford, Raymond Cattell, Philip E. Vernon i George A. Ferguson.

***

Tyle miała nam do zaproponowania psychologia, a teraz czas na… informatykę. Na szczególną uwagę zasługuje tu koncepcja definiowania inteligencji poprzez kompresję informacji. Chodzi o zdolność do znajdowania wzorów, regularności i struktur w danych tak, by reprezentować je jak najzwięźlej, bez utraty (lub przy minimalnej utracie) informacji.

W teorii informacji funkcjonuje coś takiego jak złożoność Kołmogorowa. To długość najkrótszego możliwego programu (komputerowego) służącego do wygenerowania zadanego ciągu znaków. Brzmi strasznie, wiem, już tłumaczę. Rozwinięcie dziesiętne liczby pi (3,1415926535…) jest nieskończonym i nieokresowym ciągiem cyfr (znaków), ponieważ pi jest liczbą niewymierną. Potrzebowalibyśmy nieskończonej kartki papieru, żeby je zapisać.

Istnieją jednak bardzo proste programy, takie zawierające zaledwie kilka czy kilkanaście linijek kodu, które są w stanie wygenerować dowolną liczbę cyfr jej rozwinięcia. A zatem rozwinięcie dziesiętne liczby pi, choć nieskończone, ma bardzo niską złożoność Kołmogorowa. Znaleźliśmy bowiem krótki przepis (algorytm), który tłumaczy niekończące się dane w maksymalnie zwięzły sposób. Innymi słowy: skompresowaliśmy rozwinięcie dziesiętne liczby pi do kilkunastu linijek programu komputerowego.

Jak łatwo się domyślić, dane o dużej regularności lub takie, które da się wygenerować za pomocą prostego algorytmu, mają krótką reprezentację i dają się świetnie kompresować. Dane losowe i chaotyczne — wręcz przeciwnie.

Koncepcję definiowania inteligencji poprzez kompresję informacji można by podsumować następującym zdaniem: im lepiej rozumiesz zasady rządzące światem, tym lepiej umiesz przewidzieć, co się wydarzy, tym mniej zaskakują Cię nowe dane i tym skuteczniej potrafisz je kompresować.

Pewną trudność w przyswojeniu sobie tej koncepcji może stanowić fakt, że dla wielu z nas kompresja to po prostu zmniejszenie rozmiaru pliku, bez wnikania w to, jakie mechanizmy stoją za tym procesem. Warto jednak spojrzeć na kompresję jak na zdolność przewidywania kolejnych porcji danych w oparciu o wyekstrahowane uprzednio wzorce.

Wyobraźmy sobie, że mamy pewien bardzo długi ciąg liczbowy, w którym co jakiś czas trafia się losowe zaburzenie. Jeśli umiemy przewidzieć jego kolejne elementy, pomijając zaburzenia, to znaczy, że znamy wzorce rządzące wcześniejszymi fragmentami. A skoro znamy wzorce, to potrafimy te dane skompresować, zapisując tylko różnice między naszym przewidywaniem a rzeczywistością. Innymi słowy: zamiast wypisywania całego długiego ciągu, podajemy przepis na wygenerowanie jego kolejnych elementów i uzupełniamy go o listę zaburzeń. Jeśli ten drugi zapis jest krótszy od pierwszego, to znaczy, że dokonaliśmy skutecznej kompresji.

W analogiczny sposób możemy spojrzeć na duże modele językowe, takie jak GPT, Gemini czy Claude, które dla zadanej sekwencji słów (tokenów) uczą się przewidywać jej kolejny element, kompresując tym samym język i wiedzę. Nie dysponujemy niestety oficjalnymi liczbami, które umożliwiłyby nam wyznaczenie współczynnika tej kompresji, ale z przecieków i spekulacji dowiadujemy się, że same wagi wyuczonego modelu GPT-4 mogłyby zajmować kilka terabajtów (1,8 biliona parametrów razy 4 bajty), zaś szacunki w przypadku zbioru treningowego sięgają nawet 1 petabajta, czyli 1000 terabajtów. To oznaczałoby 1000-krotną kompresję. Całkiem nieźle!

Niektórzy lubią stosować następującą analogię: dziecko uczy się języka poprzez rozpoznawanie regularności (gramatyki, znaczenia, kontekstu), by nie musieć zapamiętywać każdego zdania osobno. Kompresuje język. Podobnie robią duże modele językowe — uczą się statystycznych zależności między słowami, by przewidywać kolejne. Kompresują język. Czy to oznacza, że dzieci i duże modele językowe są równie inteligentne? Moim zdaniem nie, ale — jak w przypadku wielu rzuconych do tej pory kotwic — jeszcze do tego wrócimy.

Żeby móc na serio rozmawiać o inteligencji kompresyjnej, należy sięgnąć do prac Jürgena Schmidhubera i Marcusa Huttera, których uważa się za ojców tej koncepcji. Jeśli jakiś system potrafi wykrywać wzorce, przewidywać kolejne elementy danych oraz generować ich zwięzłą reprezentację (model świata), to spełnia definicję inteligencji w tym sensie.

Duże modele językowe robią to wszystko. Uczą się statystycznych regularności w ogromnych zbiorach danych tekstowych, optymalizują się do przewidywania kolejnego słowa (tokena) oraz tworzą wewnętrzne reprezentacje świata, które

pozwalają im skutecznie kompresować dane. Modele językowe są zatem potężnymi kompresorami semantycznymi — kompresują znaczenie, by generować treści, które zachowują sens. Mówiąc językiem teorii informacji: duże modele językowe minimalizują entropię warunkową między kolejnymi tokenami. A mówiąc językiem social mediów: Jürgen Schmidhuber lubi to!

Ale nie do końca. Ta beczka miodu zasługuje na solidną łyżkę dziegciu. Zdaniem Schmidhubera prawdziwie inteligentny system stale szuka nowych sposobów ulepszenia kompresji przez aktywne odkrywanie świata. Duże modele językowe tego nie potrafią. Są statyczne i pasywne, trenowane raz, *offline*, a potem nie zmieniają już swoich parametrów. Nie mają celu ani motywacji, nie prowadzą aktywnego uczenia się przez eksplorację, nie testują hipotez o świecie, co oznacza, że nie sprawdzają, czy ich kompresja jest dobra w praktyce.

Jeśli duży model językowy jest statycznym kompresorem języka, to prawdziwie inteligentny system byłby dynamicznym kompresorem rzeczywistości.

A jeśli nawet zdefiniujemy inteligencję jako zdolność do kompresji informacji, to komu tak naprawdę należy się order? Modelowi, który nie ma pojęcia, że cokolwiek kompresuje, który nie ma koncepcji celu poza minimalizacją błędu zadanej funkcji straty i który nie zmienia swojej architektury ani reguł uczenia? Czy raczej człowiekowi, który definiuje architekturę, określa cel, dobiera dane oraz steruje procesem uczenia?

W przypadku algorytmu kompresji ZIP nie mamy takich wątpliwości, prawda? Nie interesuje nas sama zdolność do kompresji, ale źródło tej zdolności. Czy nie powinniśmy zatem stwierdzić, że duży model językowy nie jest inteligentny sam w sobie, ale świeci światłem odbitym inteligencji kompresyjnej człowieka?

***

Ten rozdział nie byłby kompletny, gdybym nie poruszył w nim jeszcze jednego tematu, który czai się gdzieś na granicy naszego poznania. Czy inteligencja wymaga świadomości lub wolnej woli? Czy możemy nazwać inteligentnym coś, co wykonuje jedynie polecenia operatora, bez jakichkolwiek własnych intencji czy motywacji?

Intuicyjnie wolną wolę definiujemy jako zdolność do świadomego, autonomicznego wyboru działania, który wynika z własnych intencji, a nie z przymusu, oraz który mógłby być inny, gdybyśmy tak zdecydowali.

Problem w tym, że nie wiemy, na ile nasza wolna wola jest rzeczywiście wolna, a na ile nasze wybory są nam narzucone przez prawa przyrody. Zdaniem deterministów wolna wola to iluzja, subiektywne poczucie decyzji. Wszystko, co robimy, ma swoje przyczyny w biologii, psychologii i fizyce, a nasze wybory są tylko kolejnym ogniwem łańcucha przyczyn, gdyż wynikają wprost z wcześniejszych stanów naszego mózgu.

Neurobiologia zdaje się potwierdzać to stanowisko. Nasz mózg wie, co zrobimy, na moment przed świadomą decyzją. Badał to m.in. Benjamin Libet w latach 80. ubiegłego wieku. Uczestnicy eksperymentu mieli nacisnąć przycisk, kiedy zechcą, i wskazać moment, w którym poczuli, że podjęli decyzję. Badanie elektroencefalografem (EEG) wykazało, że aktywność mózgu pojawiała się ok. 350 ms wcześniej, nim uczestnicy świadomie podjęli swoje decyzje.

Późniejsze badania Haggarda, Soona, Haynesa i innych potwierdziły te wnioski również dla bardziej złożonych decyzji dzięki zastosowaniu funkcjonalnego rezonansu magnetycznego. Obserwacja aktywności w korze przedczołowej, zakręcie obręczy czy korze motorycznej pozwala na przewidzenie wyboru (np. użycia prawej lub lewej ręki) nawet na kilka sekund przed świadomą decyzją. A zatem wszystko wskazuje na to, że świadomość nie inicjuje decyzji, a jedynie ją rejestruje.

I pewnie tu moglibyśmy się zatrzymać, przyjmując stanowisko deterministów, gdyby Libet podczas swoich badań nie zauważył, że człowiek w ostatniej chwili może świadomie powstrzymać się od działania, które w sposób nieświadomy zostało zainicjowane przez jego mózg. W kontrze do *free will*, angielskiego określenia na wolną wolę, swoją koncepcję nazwał *free won't*, co moglibyśmy przetłumaczyć jako wolną wolę powstrzymania.

Perspektywa psychologiczna sugeruje również, że większość naszych wyborów jest intuicyjna i automatyczna. Nasza świadomość racjonalizuje te decyzje po fakcie, tworząc narracje, które dają nam poczucie sprawczości. Dzięki temu mamy subiektywne wrażenie wolności, nawet jeśli proces był nieświadomy. Kluczowe jest tu jednak słowo „większość".

Daniel Kahneman, guru psychologii i laureat Nagrody Nobla z 2002 roku za przełomowe badania nad ludzkimi osądami i decyzjami, zaproponował model dwóch systemów myślenia.

System 1 to myślenie szybkie — nieświadome, automatyczne i intuicyjne. Idziemy sobie ciemną ulicą, widzimy grupę głośno zachowujących się mężczyzn, przechodzimy na drugą stronę. Nie zastanawiamy się nad tym, nasza reakcja jest automatyczna. Tak wygląda absolutna większość naszych działań. Przetwarzamy ogromne ilości informacji w bardzo krótkim czasie przy minimalnym nakładzie energetycznym, ale płacimy za to błędami poznawczymi, takimi jak efekt potwierdzenia czy zakotwiczenia.

System 2 to myślenie wolne — świadome, intencjonalne i analityczne. Pozwala na kontrolowanie impulsów i rozpoznawanie własnych błędów poznawczych. Czytamy sensacyjny nagłówek, który wspiera nasze poglądy, ale zatrzymujemy się, korygujemy automatyczne sądy, sprawdzamy źródło, analizujemy i wyciągamy własne wnioski. Myślenie wolne wymaga znacznie więcej energii i zasobów poznawczych. Daje poczucie kontroli, ale jest uciążliwe i męczące.

Kahneman sugeruje, że w naszym codziennym życiu oba systemy współdziałają w podejmowaniu decyzji. Przez większość czasu jedziemy na autopilocie, a system 1 pilnuje, żebyśmy tę jazdę przeżyli i za bardzo się nie poobijali. Od czasu do czasu trafiamy jednak na złożony problem, który wymaga naszej uważności i wtedy kontrolę przejmuje system 2.

I o ile dominacja systemu 1 w naszym codziennym życiu rzeczywiście mogłaby wskazywać na brak wolnej woli, o tyle możliwość powstrzymania się od przyzwyczajeń i impulsów, rozważanie konsekwencji i alternatyw, a także świadomość i zdolność do refleksji związana z systemem 2 mogą stanowić silne argumenty za jej istnieniem.

Nie wydaje się jednak, byśmy mogli to jednoznacznie rozstrzygnąć. Być może wolna wola nie istnieje, a to, że wydaje się nam, że ją mamy, jest tylko sprytnym ewolucyjnym wynalazkiem, potrzebnym wyłącznie do tego, byśmy mogli sprawnie funkcjonować w społeczeństwie. Nie sposób przy tym wykluczyć, nie wiedząc,

czym dokładnie jest świadomość, że istnieje jakiś niezależny moment wyboru, który nie jest w pełni zdeterminowany przez biologię czy fizykę.

***

Muszę się jeszcze wytłumaczyć ze stwierdzenia, że nie do końca wiemy, czym jest świadomość. Wikipedia podpowiada, że to podstawowy i fundamentalny stan psychiczny, w którym jednostka zdaje sobie sprawę ze zjawisk wewnętrznych, takich jak własne procesy myślowe, oraz zjawisk zachodzących w środowisku zewnętrznym, na które jest w stanie reagować.

Zdaje sobie sprawę, czyli co? Uświadamia je sobie? Czujemy chyba przez skórę, że na własne życzenie pakujemy się w niezłe bagno. Świadomość jest zjawiskiem subiektywnym, wewnętrznym, nie możemy zobaczyć jej z zewnątrz ani zmierzyć jak temperaturę czy masę. Tak na dobrą sprawę jedyna świadomość, której możemy być pewni, to nasza własna. Czy ja jestem świadomy? Z pewnością, ale czy to samo mogę powiedzieć o kimkolwiek innym?

Możemy obserwować aktywność mózgu, ale nie możemy zajrzeć w czyjeś przeżycie. Na styku subiektywnych doświadczeń i obiektywnych opisów procesów fizycznych w mózgu pojawia się tzw. luka wyjaśniająca (ang. *explanatory gap*) — termin, który w 1983 roku ukuł filozof Joseph Levine. Istnieje bowiem ogromna przepaść między obserwacją, że fotoreceptory w naszych oczach reagują na falę elektromagnetyczną o długości 700 nm, sygnały biegną przez ciało kolankowate boczne i aktywuje się kora V4 w płacie potylicznym a subiektywnym doświadczeniem widzenia czerwieni.

Luka wyjaśniająca stanowi rdzeń tego, co inny filozof, David Chalmers, w 1995 roku nazwał twardym problemem świadomości. Zacznijmy jednak od problemów miękkich, co wcale nie znaczy, że łatwych do rozwiązania. Jak mózg rozpoznaje twarze? Jak rozróżnia dźwięki? Jak planuje ruchy? Jak skupia uwagę na konkretnym bodźcu? Odpowiedzi na te pytania możemy znaleźć w ramach nauki empirycznej, poprzez obserwację i eksperyment.

Twardy problem to pytanie o to, dlaczego i w jaki sposób procesy fizyczne w mózgu w ogóle rodzą subiektywne doświadczenia? To pytanie o tzw. *qualia*, czyli

wewnętrzny aspekt przeżywania. Możliwe, że nigdy nie poznamy na nie odpowiedzi. Nie ma bowiem obiektywnego testu, który mógłby wykazać istnienie świadomości, ani empirycznej teorii subiektywnego doświadczenia. Te dwa porządki po prostu nie chcą ze sobą rozmawiać.

Czy to oznacza, że nie możemy powiedzieć absolutnie nic na temat świadomości? Aż tak źle nie jest, choć sam proces pozyskiwania wiedzy w tym zakresie przypomina nieco proces poszlakowy w sądzie, gdzie zamiast przedstawienia bezpośrednich dowodów winy oskarżonego wyrok opiera się na wnioskach wyciągniętych z łańcucha wielu różnych poszlak, które razem tworzą logiczny i spójny ciąg zdarzeń prowadzących do pewności w sprawstwie.

Wiemy, że świadomość istnieje i ma charakter subiektywny, choć nie wiemy, czym właściwie jest i skąd się bierze. Wiemy, że jest ściśle związana z funkcjonowaniem układu nerwowego, i choć nie wiemy, w jaki sposób, to wiele wskazuje na to, że bez niego po prostu nie występuje. Wiemy, że ma strukturę i treść (np. emocje, myśli, wspomnienia), które można badać, choć nie wiemy, dlaczego akurat taką strukturę i taką treść. Wiemy, że pełni funkcje poznawcze (np. umożliwia refleksję nad sobą i światem, pośredniczy w podejmowaniu złożonych decyzji), choć nie wiemy, czy to te funkcje implikują świadomość, czy wręcz przeciwnie — są jej efektem.

Wiemy, że w naszych mózgach próżno szukać jednego centrum świadomości. Wydaje się, że jest to tzw. własność emergentna współdziałania wielu obszarów. Badania prowadzone przy użyciu funkcjonalnego rezonansu magnetycznego wskazują, że bodźce, które postrzegamy świadomie, wywołują szeroko rozproszoną aktywację, zaś nieświadome tylko lokalną. Do podobnych wniosków prowadzą badania nad snem głębokim i narkozą — aktywność mózgu nie zanika, neurony funkcjonują lokalnie, ale tracimy globalną integrację sieci, co wiąże się z utratą świadomości.

Fizykalizm redukcyjny zakłada, że… Dość! Stop! Już wystarczy! Czyż to nie zabawne, że choć o świadomości wiemy tak zaskakująco niewiele, to możemy pisać o niej bez końca? A może właśnie dlatego…? Czas na wielkie podsumowanie!

***

Zdaje się, że już na samym początku rozdziału niechcący zdradziłem puentę: nie ma jednej, spójnej i powszechnie akceptowanej definicji inteligencji. Śledztwo, które prowadziliśmy, wcale nie miało nas do niej doprowadzić, a jedynie wykazać, dlaczego nie jesteśmy — i być może nigdy nie będziemy — w stanie jej opracować. Perspektywa psychologa, neurobiologa, filozofa i informatyka, choć może zawierać części wspólne, skupia się na zupełnie innych aspektach inteligencji.

Gdy pytamy, czy inteligencja wymaga świadomości, zwolennik funkcjonalizmu odpowie, że nie — inteligencja to przetwarzanie informacji, które może zachodzić bez przeżywania. Gdy o to samo zapytamy fenomenologa, dowiemy się, że prawdziwa inteligencja, czyli zdolność do rozumienia, elastycznego myślenia i samoregulacji, nie może istnieć bez pewnej formy samoświadomości. Z kolei badacz ludzkiego mózgu, chcąc zachować daleko posuniętą powściągliwość, stwierdzi, że świadomość to nie warunek inteligencji, ale narzędzie dla bardziej złożonych form inteligencji refleksyjnej. Żaden z nich nas nie okłamie. Każdy udzieli najlepszej możliwej odpowiedzi w oparciu o swoją wiedzę i doświadczenie.

Spostrzegawczy obserwator bez trudu zauważy, że tam, gdzie zaczyna się niepewność, pojawiają się przymiotniki. Stąd mowa o inteligencji prawdziwej, refleksyjnej, obliczeniowej, operacyjnej i Bóg wie jakiej jeszcze. Aż kusi, by przypomnieć żart o krześle i krześle elektrycznym.

Gdy mówimy o generatywnej sztucznej inteligencji i dużych modelach językowych, bez trudu znajdziemy przymiotniki i definicje, które pozwolą nam stwierdzić, że z całą pewnością jest to jakaś forma inteligencji. Nie trzeba się nawet szczególnie gimnastykować, wystarczy sięgnąć po definicję inteligencji kompresyjnej i nie zadawać niewygodnych pytań o to, kto zaprojektował algorytm uczący. I z tym w zasadzie nie ma co polemizować.

Natomiast to, co budzi mój autentyczny sprzeciw, to twierdzenie, że sztuczna inteligencja myśli lub rozumie. Bo o ile jestem w stanie przełknąć funkcjonalną interpretację słowa „inteligencja", o tyle myślenie i rozumienie mają dla mnie zupełnie inny — rzekłbym: ontologiczny — ciężar gatunkowy. Nie chodzi o to, że nie podoba mi się ich potoczne czy funkcjonalne użycie, ale o to, że zbyt często w ich kontekście słyszałem niedopowiedziane, choć domniemane „jak człowiek". I tu pojawia się zasadniczy problem.

Sztuczna inteligencja nie myśli jak człowiek. Jeżeli czegokolwiek możemy być pewni, to właśnie tego. Nie mamy żadnych podstaw, by przypisywać jej świadome przeżycia, własne znaczenia, perspektywę pierwszoosobową, introspekcję, intencjonalność, emocje, motywację, potrzeby, pragnienia, cele, jaźń, narracje wewnętrzne, ucieleśnione doświadczenia czy podmiotowość.

Czy to się kiedykolwiek zmieni? Czy zacznie myśleć jak człowiek? Czy uzyska świadomość i wolną wolę? Czy da się odtworzyć ludzką inteligencję w komputerze? Obawiam się, że to nie koniec, a ledwie początek naszego wspólnego śledztwa. Bo choć na te pytania współczesna nauka nie zna jeszcze odpowiedzi, to przy odrobinie szczęścia, dociekliwości i samozaparcia znajdziemy wystarczająco dużo poszlak, by zbudować przekonujący obraz nadchodzącej przyszłości…

Zdarzyło Ci się usłyszeć, że sztuczna inteligencja już dziś myśli jak człowiek? Że wystarczy więcej danych, większy model i trochę czasu, a powstanie AGI — maszyna zdolna rozumieć, tworzyć i decydować, przewyższająca nas niemal we wszystkim, co robimy? A może dotarła do Ciebie jeszcze bardziej niepokojąca opowieść: że na własnych oczach hodujemy zupełnie nowy gatunek — obcy, szybszy i doskonalszy od nas — który w najlepszym razie podporządkuje sobie ludzi, a w najgorszym doprowadzi do ich zagłady? Skoro AI teraz robi rzeczy, które jeszcze niedawno wydawały się niemożliwe, to co się stanie za pięć czy dziesięć lat?

W eseju popularnonaukowym, poświęconym jednemu z najważniejszych sporów naszych czasów — o to, czym naprawdę jest inteligencja i czy da się ją odtworzyć w maszynie — autor prowadzi czytelników przez intelektualny labirynt, w którym spotykają się informatyka, filozofia umysłu, neurobiologia, psychologia rozwoju i zdrowy sceptycyzm wobec technologicznych obietnic. Od chińskiego pokoju i filozoficznych zombie, przez granice systemów formalnych i uczenia statystycznego, aż po poznanie ucieleśnione, emergencję, metapoznanie i zawodność benchmarków — ta książka pokazuje, że pytanie o AI nie dotyczy wyłącznie maszyn. Dotyczy przede wszystkim nas: tego, czym jest rozumienie, skąd się bierze sens i dlaczego sprawność w rozwiązywaniu zadań nie musi jeszcze oznaczać inteligencji.

To nie jest pamflet przeciw technologii ani kolejna pochwała cyfrowej przyszłości. To rzecz dla tych, którzy nie chcą już bezrefleksyjnie powtarzać obietnic Doliny Krzemowej ani katastroficznych proroctw o końcu człowieka. Jeśli chcesz zrozumieć, czym AI naprawdę jest, a czym nie jest, i dlaczego pytanie o maszynową inteligencję okazuje się w gruncie rzeczy pytaniem o nas samych — wejdź w ten spór, zanim pozwolisz maszynom zdefiniować myślenie za Ciebie.

Helion

helion.pl

HELION S.A.
ul. Kościuszki 1c
44-100 Gliwice
tel.: 32 230 98 63
helion@helion.pl

KOD KORZYŚCI
Sięgnij po więcej! ▶

ISBN 978-83-289-4195-3

9 788328 941953

Cena: 59,90 zł