

Epistemiczne bankructwo

czyli ile możemy dowiedzieć się od
Wielkich Modeli Językowych

Cezary Szmidke

Wstęp

Korzystając z technologii cyfrowych w 2026 roku nie sposób uniknąć LLM-ów, czyli Wielkich Modeli Językowych (ang. *Large Language Models*). Narzędzia te są nie tyle reklamowane, co wręcz forsowane przez technologicznych gigantów. Spotykamy się z nimi podczas edycji własnych tekstów, wiele treści w Internecie – a coraz częściej także tych pojawiających się w druku – powstaje przy ich użyciu. Do ich stosowania zachęca się pracowników na przeróżnych stanowiskach, a także studentów uczelni. Zastąpiły w znacznej mierze rozmowy z konsultantami, a także tradycyjne wyszukiwarki internetowe.

Nie jest to jednak, jak niektórzy utrzymują, „tylko narzędzie”: w przeciwieństwie do młotka czy nożyczek ich funkcjonowanie jest nierozzerwalnie związane kwestiami własności intelektualnej (modele te były i są budowane w oparciu o teksty, których twórcy nie byli nigdy pytani o zgodę) i autorstwa wygenerowanych treści, z geopolityką i wielkim kapitałem (serwery oferujące możliwość korzystania z LLM-ów są prywatne i znajdują się często w innym kraju, niż użytkownik), z prywatnością (co dzieje się z pytaniami i odpowiedziami?), z ogromnym wpływem na rynek pracy, na środowisko, na rozwój – czy może raczej zanik – kompetencji u użytkowników, zwłaszcza tych młodych.

To wszystko ważne, ale w publicznym dyskursie bagatelizowane i marginalizowane aspekty tej technologii. Jako filozofa frapuje mnie szczególnie inna, jeszcze rzadziej poruszana kwestia: na ile można polegać na treściach generowanych przez LLM-y? Nie ulega wątpliwości, że wiele osób podejmuje decyzje na podstawie wyników działania tych algorytmów, traktując te wyniki jako źródło wiedzy i rozróżniając „prawdziwe” odpowiedzi i „halucynacje”. Uważam, że rozróżnienie to jest nieuzasadnione, a modele językowe, pomimo rozwoju technologicznego, są i pozostaną fundamentalnie bezużyteczne dla każdego, kto ceni prawdę – czyli, mam nadzieję, przynajmniej dla miłujących mądrość. Stanowisko to wyda się być wcale nie radykalnym, może wręcz oczywistym, kiedy skonfrontujemy mechanizm działania LLM-ów z pojęciami prawdy i wiedzy.

Duch w maszynie

Modele językowe, zgodnie ze swoją nazwą, modelują użycie języka. Interakcja z nimi odbywa się przy użyciu języka naturalnego, w związku z czym łatwo o personifikację tych narzędzi, tym bardziej, skoro mechanizm ich działania jest trudny do zrozumienia dla osób nieorientowanych w uczeniu maszynowym, specjalizacji w ramach techniki informatycznej. Większość użytkowników traktuje je więc z konieczności jak „czarne skrzynki”: podajemy dane wejściowe, dzieje się coś tajemniczego, otrzymujemy dane wyjściowe. Te tajemnicze działania są w istocie inteligibilne, mało kto jednak ma dość ciekawości oraz czasu, aby je zgłębić. Na potrzeby dalszej argumentacji postaram się przybliżyć zasadę działania tychże modeli możliwie prosto i bez wdawania się w kwestie techniczne.

Wszystkie LLM-y działają na komputerach o takiej samej architekturze, jak te używane na co dzień przez miliardy ludzi na świecie. To znaczy, że wykonują sekwencyjnie operacje syntaktyczne. Mnóstwo operacji w ramach skomplikowanych algorytmów, ale fundamentalnie są to następujące po sobie obliczenia.

Powszechnie używane modele językowe nazywa się wielkimi, ponieważ do ich wytworzenia – wyszkolenia – potrzeba ogromnych ilości materiału: książek, artykułów, komentarzy pozostawionych przez użytkowników witryn internetowych, zapisów wszelkiego rodzaju językowej komunikacji w modelowanym języku. Szkolenie polega na analizie statystycznej prawdopodobieństwa współwystępowania i następowania po sobie słów, morfemów, znaków interpunkcyjnych itd., łącznie nazywanych tokenami. Na gotowy model składają się miliony, a nawet miliardy tak zwanych parametrów, czyli wag wyrażających wspomniane prawdopodobieństwa. Przeanalizowanie olbrzymiej ilości danych i wyrażenie relacji między tokenami w postaci parametrów wymaga stosownie wielkich zasobów, zarówno sprzętowych jak i energetycznych, w związku z czym jedynie największe instytucje mogą pozwolić sobie na podobne przedsięwzięcie. Można w związku z tym dostrzec pewne analogie względem kontroli środków produkcji w efekcie rewolucji przemysłowej w XIX wieku, ale na potrzeby obecnych rozważań ograniczymy się do kwestii epistemicznych.

Wyszkolony model jest gotowy do inferencji, czyli do interakcji poprzez np. zadanie pytania. Wpisane przez użytkownika pytanie, rozbite na tokeny (słowa, morfemy itd.), staje się częścią kontekstu, a model językowy – stosując parametry w połączeniu z liczbami pseudolosowymi¹ – dobiera kolejne tokeny, które na podstawie materiału szkoleniowego

1 Komputery są maszynami deterministycznymi, stąd nie są w stanie generować liczb losowych podobnie skutecznie, jak choćby rzut kością do gry.

powinny prawdopodobnie znaleźć się na kolejnym miejscu. To, czego użytkownik nie widzi, to część kontekstu wprowadzona przez twórców modelu, która choć ukryta, również jest brana pod uwagę podczas inferencji. Raz wyszkolony model nie zmienia się, nie uczy się niczego nowego (w rozumieniu uczenia maszynowego, nie takiej nauki, jakiej podjąć mogą się np. ludzie): podczas dłuższej interakcji rozbudowuje swój kontekst i bierze pod uwagę wcześniejsze pytania i odpowiedzi, ale każde ponowne uruchomienie ma ten sam punkt wyjścia.

Aby każdy kolejny etap tego procesu był efektywny, potrzeba nie lada inwencji w stworzeniu instrukcji dla komputera, ale niezależnie od producenta i kolejnej wersji danego modelu, powyższy schemat pozostaje ten sam. Gigantyczny zbiór zapisów użycia języka naturalnego zostaje ujęty w tokeny reprezentujące słowa, morfemy, interpunkcję itd., relacje między nimi są wyrażone w postaci parametrów, użytkownik w ramach inferencji dodaje swoje zapytanie do kontekstu, a model dodaje kolejne tokeny zgodnie z prawdopodobieństwem ich wystąpienia.

Niezaprzeczalnie zdumiewającym jest fakt, iż dzięki analizie syntaktyki języka naturalnego algorytm jest w stanie w tak przekonujący sposób symulować użycie tegoż języka. Czy wystarczy to jednak, abyśmy mogli wynik działania takiego algorytmu uznać za prawdziwy, albo czegoś się z niego dowiedzieć?

Cóż to jest prawda?

Prawda, jej miejsce w ontologii i uchwytność, interesuje głównie filozofów. Naukowcy i badacze, zwłaszcza zajmujący się naukami przyrodniczymi, przede wszystkim potrzebują jej dla uzasadnienia wiedzy, przyjmując zwykle *implicite* że prawdziwe są fakty, do których starają się dopasować jak najlepszą teorię. Dla wszystkich innych (a kiedy nie skupiają się właśnie nad swoją dziedziną, dla filozofów i naukowców także) najważniejsza wydaje się jednak pragmatyka: jeśli *wiem* o której godzinie odjeżdża pociąg, nie spóźnię się na niego. Jeśli majster mówi *prawdę*, wykona tylko potrzebne naprawy. Jeśli *nie wiem* jak rozliczyć podatki, zapytam księgowej. Jeśli plotka okaże się *nieprawdą*, mogę znaleźć się w niezręcznej sytuacji. Mało kto w podobnych sytuacjach zastanawia się nad identycznością, koherencją czy korespondencją. Niezmiennie jednak używamy heurystyk odnoszących się do doświadczenia, autorytetu czy sylogizmów. Uzupełniamy naszą *wiedzę* sięgając po dodatkowe informacje zakładając, że są one *prawdziwe*. Pytamy specjalistów, przeglądamy literaturę, dedukujemy, a wielu z nas coraz częściej opiera się na odpowiedzi wygenerowanej przez LLM.

Systematyczny przegląd różnych koncepcji prawdy, choćby tych opracowanych w XX wieku i później (Candlish & Damnjanovic, 2007), wykracza poza zakres tej pracy. Aby sprawdzić, czy wynik działania algorytmów modeli językowych możemy traktować jako prawdziwy i czy możemy uzyskać dzięki niemu wiedzę, wystarczy nam jednak skonfrontować mechanizm działania tego narzędzia z wybranymi cechami różnorodnych poglądów na prawdę i wiedzę.

Prawda i język

Mogło by się wydawać, że wobec prawdy jak traktował ją Fryderyk Nietzsche, wszelkie zarzuty, jakie moglibyśmy sformułować wobec LLM-ów, nie będą miały większego znaczenia. To, co nazywamy prawdą, i tak jest według Nietzschego złudą (Nietzsche, 1873, s. 170). Podkreśla on jednak, że o ile ten paradoks – prawda jako złuda – bynajmniej ludziom nie przeszkadza, obawiają się oni „przykrych następstw pewnego rodzaju złudzeń”, a pożądają „miłych, wspierających życie konsekwencji” (*ibidem*, s. 162). Nie jest to kryterium istotowe, ale pragmatyczne i do tego subiektywne. Dopóki więc model językowy zadowala użytkownika dając prędkie odpowiedzi, czy – w ramach tej dość szczególnej koncepcji – można je uznać za prawdziwe? Nie sądzę, ponieważ prawda-złuda jest też „ruchliwą armią metafor, metonimii (...), ludzkich stosunków” (*ibidem*, 164). O ile ktoś mógłby utrzymywać, że te metafory i stosunki nie zostają utracone w wyniku tokenizacji, analizy statystycznej i inferencji, w modelu językowym nie ma żadnej ruchliwości i zmienności. Tego rodzaju zamrożenie stoi w sprzeczności z dynamicznym, witalistycznym charakterem filozofii Nietzschego. Choć według niego „prawda jest człowiekowi obojętna”, jej szukanie polega na „właściwym podporządkowaniu poszczególnych przypadków danemu pojęciu”, które to pojęcie jest naszym działaniem (Nietzsche, 1889, s. 208). W przypadku LLM-ów to algorytm działa, a nie my. „Właściwość przyporządkowania” nie zostaje ściślej określona, ale nie sądzę, aby brak transparentności w doborze materiałów, które zostały użyte do wyszkolenia modelu oraz niejawność części kontekstu podczas inferencji były w tym sensie właściwe. Wyglądają raczej, używając terminów Nietzschego, na działanie typowe dla kapłanów. Zanim zgodzimy się, że prędkość uzyskania odpowiedzi z modelu językowego jest dla użytkownika „miła i wspierająca życie”, powinniśmy zwrócić także uwagę na dobrze już znane konsekwencje polegania na tych narzędziach, szczególnie na utratę umiejętności związanych z działaniami, w których pozwalamy się wyręczać.

Znacznie więcej uwagi zagadnieniom prawdy i wiedzy poświęcił Ludwik Wittgenstein w kontekście gier językowych. W ramach tej koncepcji nie musimy polegać na domysłach, bo choć Wittgenstein niewiele pisał o automatach, jego refleksje na temat prawdy i wiedzy możemy odnieść do LLM-ów i naszych z nimi interakcji. Na pierwszy rzut oka, skoro za słowem nie stoi jego istota jako przedmiot drugiego rzędu i decydujący jest kontekst użycia słowa, wystarczająco sprawny i przekonujący model językowy mógłby zwieść kogoś jako partner gry językowej. Wittgenstein wprost jednak mówi o zasadniczym znaczeniu przejrzystości, która umożliwia zrozumienie polegające na dostrzeżeniu związków między słowami (Wittgenstein, 1958, s. 75). Przejrzystość i zrozumienie są odnoszone to do stanów psychicznych, to do zjawisk w mózgu, czy nie mogą więc również być reprezentowane przez parametry modelu językowego? Osobiście nie zgodziłbym się z takim stanowiskiem, ale nawet gdyby na potrzeby dyskusji przyjąć, że modelom przysługuje przejrzystość i zrozumienie, skoro są w stanie – jak się zdaje – stosować słowa we właściwy sposób, okaże się, że pomimo powierzchownych podobieństw nie używają tego samego języka, nie grają w tę samą grę, ponieważ „układem odniesienia jest tu wspólny ludziom sposób działania” (*ibidem*, s. 121). Dwa zdania o tej samej formie graficznej nie są ze sobą identyczne, skoro powstały w ramach różnych reguł: zdania z zupełnie obcego języka, czyli z innej gry, nie możemy zrozumieć i nie możemy dowiedzieć się tego, co zdawałoby się ono oznaczać, gdyby było zdaniem naszego języka, naszej gry. Wittgenstein ilustruje to przekonująco przykładem dwóch osób, A i B, komunikujących się za pośrednictwem pisma. Osoba A zapisuje liczby naturalne tworzące pewien ciąg arytmetyczny. Jeśli B rozpozna zasadę ciągu, może powiedzieć: „teraz rozumiem”, „teraz wiem, jak dalej”, „teraz mogę kontynuować” (dopisywanie kolejnych liczb), a to dlatego, że kieruje się tą samą regułą, co A (*ibidem*, s. 88-91).

Jeśli wprawnemu grzybiarzowi opiszę znaleziony przeze mnie okaz grzyba i zapytam, czy jest on jadalny, na co uzyskam odpowiedź: „Tak, ten grzyb jest jadalny”, dowiem się czegoś, ponieważ ja i ekspert mamy ten sam sposób działania, rozpoznaję go jako autorytet ponieważ rozumiem, że skończył liczne kursy i nabył praktyczne doświadczenie: kierujemy się tymi samymi regułami używając w ten a nie inny sposób słów „grzyb”, „jadalny” i tak dalej. Jeśli na to samo pytanie zastosowane w inferencji modelu językowego otrzymam w rezultacie taki sam ciąg znaków: „Tak, ten grzyb jest jadalny”, niczego się nie dowiem, ponieważ nie może być tu mowy o tym samym sposobie działania i tej samej regule. Sposób działania i regułę,

która pozwoliła LLM-owi na wygenerowanie odpowiedzi możemy poznać w zupełności i nie ma wątpliwości, że są one zupełnie inne niż te używane przez ludzi.

Sens i znaczenie

Rozważania Wittgensteina wskazują na ważną intuicję: używając języka, formułując zdania i wyrażając sądy o coś nam chodzi. Systematyczne ujęcie tej intuicji możemy znaleźć u Gottloba Fregego (Frege, 1892), który rozważając różnicę między wartością poznawczą zdań $a = a$ oraz $a = b$ stosuje terminy: znaczenie, sens, obraz, myśl oraz sąd.

Znaczenie jest obiektywną częścią nazwy lub zdania, jego desygnatem. „Gwiazda Poranna” i „Gwiazda Wieczorna” mają więc to samo znaczenie. Sens to sposób, w jaki przedmiot jest dany: obserwując Wenus o brzasku lub o zmierzchu możemy użyć odpowiednio nazw o różnym sensie, ale tym samym znaczeniu. Obraz (albo przedstawienie) to subiektywne, podmiotowe przeżycie towarzyszące sensowi, nie ma on jednak związku z prawdziwością zdania, nie będziemy więc w ogóle odnosić się do różnic psychologicznych między człowiekiem a maszyną, za Fregem uważać jednak będziemy, aby znaczenia i sensu nie mylić z osobistymi doznaniem. Myśl zaś to obiektywna treść aktu myślenia, a sąd to uznanie jej prawdziwości (*ibidem*, s. 70). Jak w tej siatce pojęć odnaleźć może się wynik działania modeli językowych?

Sprawa jest nader prosta: w przypadku działania modelu językowego nie możemy w ogóle mówić o znaczeniu. Co prawda w dziedzinie uczenia maszynowego pojawia się termin „semantyka”, ale nie należy mylić go z tym samym znakiem, ale innym znaczeniem, które ma w językoznawstwie lub filozofii. Jak wszystkie operacje wykonywane przez komputer, algorytmy LLM-ów operują tylko i wyłącznie na poziomie syntaktycznym. Jeśli próbowalibyśmy uzyskać definicję słowa (tokenu) w ramach modelu językowego, sprowadzałaby się ona do prawdopodobieństwa współwystąpienia z innymi słowami (tokenami), nie zaś do znaczenia jak rozumiane przez Fregego.

Przyjrzyjmy się temu na przykładzie zdania: „Śnieg jest biały, a trawa jest zielona”. Każde z tych słów ma dla nas jasne znaczenie. To samo znaczenie, ale inny sens moglibyśmy wyrazić zdaniem: „Śnieg odbija pełne spektrum światła widzialnego, a trawa odbija fale świetlne o długości od 495 do 570 nanometrów”. Wyobrażamy sobie zapewne przedstawienie śnieżnego puchu i trawnika, w ramach sądu uznajemy myśl wyrażoną zdaniem jako prawdziwą, odwołując się do znaczeń poszczególnych elementów zdania. Na tej podstawie

możemy także stwierdzić, że dwie powyższe parafrazy (wyrażające to samo znaczenie, ale różne sensy, a więc i różne myśli) są z logicznego punktu widzenia w relacji $a = b$, co niesie ze sobą wartość poznawczą (poznajemy fizyczną charakterystykę światła o barwie zielonej).

Tymczasem model językowy, któremu w ramach inferencji podamy pytanie o określenie prawdziwości powyższego zdania najpewniej zwróci odpowiedź twierdzącą, ale nie będzie to miało związku ze znaczeniem słów „śnieg”, „biały”, „trawa”, „zielony”, „jest”, a z częstością ich współwystępowania w różnych kontekstach w materiale szkoleniowym modelu. Wynik, podobnie jak w omawianych wcześniej grach językowych, będzie graficznie równoważny, ale nie można uznać go za prawdziwy, ponieważ nie ma związku ze znaczeniem poszczególnych znaków oraz całego zdania (Candlish & Damjanovic, 2007, s. 291). Dopiero czytelnik, rozpoznając poszczególne znaki, dopasowuje do nich znaczenia, ale powinien traktować je jak dobrane losowo, bez porządkującej je myśli, którą zdają się wyrażać – ponieważ tak jest w istocie: nie są ze sobą związane logicznie, a probabilistycznie.

Wiedza przypadkowa

W ten sposób trafiamy w obszar uzasadnienia, jeden z fundamentów klasycznej definicji wiedzy – prawdziwego, uzasadnionego przekonania. Uzasadnienie wiedzy przez bezpośrednie doświadczenie ma długą tradycję (Platon, *Teajtet*, s. 106) i stanowi podstawę dla większości, jeśli nie wszystkich nauk – szczególnie przyrodniczych. Od oświecenia po neopozytywizm, odwołanie do doświadczenia było kluczowe dla empiryzmu i współczesnej metody naukowej. Zdaje się, że nawet wobec kryzysu zaufania do nauki ludzie przede wszystkim skłonni są do polegania na własnym doświadczeniu, nawet jeśli nie zawsze wystrzegą się różnych błędów poznawczych.

Stosowanie metody naukowej historycznie okazało się niebywale skuteczne w praktyce, o czym świadczy ogromny rozwój nauki, do którego się przyczyniła, zwłaszcza w ciągu ostatnich dwóch stuleci. Z filozoficznego punktu widzenia metody uzasadniania twierdzeń naukowych (i nie tylko naukowych) przedstawiają jednak liczne problemy (Motycka, 1984, s. 60), spośród których wiele jest nierozwiązywalnych, jak choćby problem indukcji. Zwykle nie prowadzi to jednak do kapitulacji, a raczej do kolejnych prób wypracowania takich sposobów uzasadniania, które będą skuteczne oraz odpowiadać będą naszym intuicjom związanym z tym, co nazywamy wiedzą i prawdą.

Potrzebę obwarowania uzasadnienia wiedzy zestawem dodatkowych wymogów unaocznili m.in. Gettier w swoich przykładach wiedzy przypadkowej, czyli przekonań, które chociaż są prawdziwe i – w formalny sposób – uzasadnione, wywołują zgrzyt, jeśli mielibyśmy nazwać je wiedzą. Sądzę, że każdy wynik działania LLM-ów należy do podobnej kategorii.

Proponenci kolejnych odsłon korespondencyjnych teorii prawdy i wiedzy w różny sposób starali się przedstawić związek przekonań, języka i faktów (Candlish & Damnjanovic, 2007, s. 288). Nawet w swoich bardzo minimalistycznych wersjach korespondencji Alston i Mackie nie twierdzą, że izomorfizm pomiędzy sądem a opisem rzeczywistości jest wystarczający, aby można było mówić o prawdziwości sądu: zdanie (*proposition*) i fakt muszą pozostawać ze sobą w relacji korespondencji poprzez wspólną treść (patrz: znaczenie u Fregego). Searle zaletę widział w nieokreśloności słowa „korespondencja”: tak, jakby chciał powiedzieć, że zdanie i fakty są ze sobą w *jakimś* związku, ale nie wiemy, jakiego rodzaju jest to związek (*ibidem*, s. 291). W przypadku tych najoszczędniejszych z punktu widzenia terminologii teoriach chyba najlepiej widać, że analiza statystyczna podczas szkolenia modelu i inferencja podczas korzystania z niego są z zupełnie innego porządku, niż ta nieostra znaczeniowo korespondencja. Na czym by nie polegała dokładnie zasada łącząca prawdziwość zdania z faktem w świecie, nie może ona przetrwać tokenizacji i parametryzacji: choćbyśmy chcieli utrzymywać, że jest to jakiegoś rodzaju mapa, jej związek z partykularnymi elementami terytorium jest nie do prześledzenia i nie może być traktowany jako metoda uzasadnienia.

Między prawdą a fikcją

Kolejną grupą ujęć prawdy i wiedzy są teorie koherencyjne, dla których kluczem jest stworzenie wewnętrznie spójnego systemu sądów powiązanych ze sobą logicznie. Model językowy jako taki nie zawiera w sobie sprzeczności, składa się jednak z parametrów opisujących odległości pomiędzy jego elementami w wielowymiarowej przestrzeni – w tym sensie jest jak najbardziej spójny. Jeśli jednak wynik inferencji, to jest zestawione ze sobą tokeny odczytamy jako słowa, morfemy itd. odczytamy jako zdania języka naturalnego, prędko odkryjemy liczne sprzeczności. Co więcej, między kolejnymi zdaniami nie ma żadnych związków logicznych: każdy kolejny token znajduje się po poprzednim jedynie w związku z parametrami wyrażającymi współwystępowanie w danym kontekście. Doprawdy, zaskakujące jest to, jak często można odnieść wrażenie, że dane wyjściowe modelu językowego są zbieżne z prawdą (nigdy: prawdziwe). Trafnie spostrzegł Nietzsche, iż

człowiek „ma niezwalczoną skłonność do ulegania złudzie, jest oniemiały ze szczęścia, gdy rapsod opowiada mu epickie baśnie jak coś prawdziwego” (Nietzsche, 1873, s. 170).

Być może ktoś znalazłby miejsce dla LLM-ów jako źródła wiedzy w kontekście teorii pragmatycznych, opierając się na niczym innym jak skuteczności dla działania i przewidywania. Oparcie się o koncepcje takie jak fikcjonalizm (Vaihinger, 1924, s. 80) pozwoli nawet pogodzić świadomość pojawiających się błędów z całym systemem wiedzy: skoro prawda jako taka i tak nie jest nam dostępna i skazani jesteśmy na poruszaniu się w ramach fikcji, dlaczego by nie zgodzić się na błąd tu i ówdzie, jeśli tylko częściej rezultat jest dla nas zadowalający?

Ile warta jest prawda

Wygląda na to, że aby usprawiedliwić traktowanie danych wyjściowych modeli językowych jako potencjalnego źródła wiedzy, to jest prawdziwych, uzasadnionych przekonań, musielibyśmy przyjąć bardzo specyficzne, niekoniecznie intuicyjne koncepcje prawdy i wiedzy. Nie jest to bynajmniej nowy dylemat: o podobnym problemie możemy przeczytać u Platona (Platon, *Gorgiasz*).

Zagadnienie stosowania LLM-ów i polegania na nich przez tak wiele osób w dużym stopniu przypomina mi rozmowę Sokratesa z mistrzem i adeptami retoryki. Tak w dialogu Platona, jak i w kontekście modeli językowych przedstawia nam się sztukę czy narzędzie, które mają być dobre – najlepsze! – do wszystkiego, lada moment zastąpić lekarzy i wszelkich innych specjalistów. Skuteczność tych metod i wielkie zyski odnotowane przez tych, którzy je stosują mają być miarą ich wielkości. Być może najbardziej jaskrawe podobieństwo to jest w pochlebstwie: tak sofista, jak i model językowy są przekonujący dlatego, że słyszymy lub czytamy rzeczy nam miłe. Przymilność modeli językowych jest dobrze znana ich użytkownikom, ale tak jak świadomość, że Kallikles jest zdolnym mówcą nie wystarczy, aby przełamać jego retoryczne chwyt, tak też nie wystarczy zauważyć, że LLM skwapliwie nam przytakuje, aby nie utwierdzać się w niekoniecznie uzasadnionym przekonaniu o własnej błyskotliwości w bieżącym temacie.

Sokrates mógłby stosować sztuki retoryczne, aby zdobyć pieniądze, odziać się w kosztowny płaszcz, zadbać o swój interes przed sądem. My dziś moglibyśmy stosować modele językowe, aby uniknąć pisania tekstów, na które nie mamy ochoty, otrzymać odpowiedź na pytanie, kiedy nie chcemy wertować książek czy kolejnych stron internetowych (wygląda na to, że

w tym porównaniu sztuki retoryczne są bardziej atrakcyjne). To wszystko można zrobić, ale ponosząc nie lada cenę: kosztem takiego działania jest prawda.

Prawdę i wiedzę, jak widzieliśmy, filozofowie ujmują na różne sposoby. Modele językowe, ze względu na swoją konstrukcję i mechanizm działania, nie są w stanie sprostać żadnej z tych koncepcji, chyba że któraś odbiega radykalnie od naszych intuicji związanych z tymi terminami. Za którąkolwiek szkołą i myślą filozoficzną nie pójdziemy, nie może być mowy o dowiadrywaniu się czegoś o świecie i o prawdziwości otrzymanych wyników. Jeśli mimo to ktoś postanawia polegać na tych narzędziach (doprawdy trudno powiedzieć, jakie może być dla niego sensowne zastosowanie) i traktować ich dane wyjściowe jako sensowne, jako prawdziwe, podejmować na ich podstawie decyzje – robi to za cenę prawdy, przedkładając ponad jej wartość szybką wygodę, łatwość, pozór wiedzy i pochlebstwo. To właśnie nazywam epistemicznym bankructwem.

Nie jest bynajmniej pewne, czy nawet największym staraniem, stroniąc od modeli językowych i innych zwodniczych dróg, możemy sięgnąć prawdy, czy też: Prawdy. Czy jest ona jednak osiągalna czy nie, w zupełności albo w części, powinna być traktowana jako niezrównana wartość, do której należy dążyć, do czego zachęca nas także Platon (*ibidem*, s. 132). Nie raz zdarzy się zbłądzić, ale nie tylko efekt – kosztowne płaszcze i tym podobne – świadczy o tym, czy działamy słusznie. Jak pisał Cynceron: „*Errare mehercule malo cum Platone... quam cum istis vera sentire*” – „na Herkulesa, lepiej mi błędzić z Platonem, niż mieć rację wraz z nimi”. Choćby i wynik otrzymany z LLM-a był zbieżny z prawdą, wolę pomylić się – ale mieć prawdę za cel.

Ceterum censeo Magnam Linguae Formam delendam esse.

Bibliografia

Candlish, S. & Damnjanovic, N. (2007). A brief history of truth. *Philosophy of Logic*. Amsterdam: Elsevier, 2007.

Frege, G. (1892). Sens i znaczenie. *Pisma semantyczne*. Tłum. Bogusław Wolniewicz. Warszawa: PWN, 2014

Motycka, A. (1984). Relatywistyczna wizja nauki. Wrocław: Ossolineum, 1984.

Nietzsche, F. (1873). O prawdzie i kłamstwie w pozamoralnym sensie. *Pisma pozostałe*. Tłum. Bogdan Baran. Kraków: inter esse, 2004.

Nietzsche, F. (1889). Fragmenty 1869 – 1889. *Pisma pozostałe*. Tłum. Bogdan Baran. Kraków: inter esse, 2004.

Platon. *Gorgiasz*. Tłum. Paweł Siwek. Warszawa: PWN, 1991.

Platon. *Teajtet*. Tłum. Władysław Witwicki. Kęty: Wydawnictwo ANTYK, 1999.

Vaihinger, H. (1924). *The Philosophy of 'As if'*. Tłum. C.K. Ogden. Londyn: Kegan Paul, Trench, Trubner & Co., 1924.

Wittgenstein, L. (1958). *Dociekania filozoficzne*. Tłum. Bogusław Wolniewicz. Warszawa: PWN, 2000