WŁODZISŁAW DUCH 15 października 2024

Sztuczna inteligencja coraz prawdziwsza. Czy właśnie zaczyna czuć i myśleć jak człowiek?

AI nic nie rozumie, nie myśli, nie jest kreatywna. Nie może być świadoma i nie ma zdolności odczuwania emocji. To bowiem potrafi tylko człowiek. Czy na pewno?

Błazen Stańczyk udowodnił, że najwięcej jest w Polsce lekarzy – pisał Józef Ignacy Kraszewski, bo prawie każdy udzielał mu porad, jak leczyć bolące zęby. Teraz każdy ma swoją opinię na temat sztucznej inteligencji. Pojawiło się wielu ekspertów, o których rok wcześniej nikt nie słyszał, nie napisali o AI żadnej publikacji, nawet nie byli na żadnej liczącej się konferencji. Zamiast argumentów padają wygłaszane z przekonaniem stwierdzenia, że to żadna inteligencja, nic nie rozumie, nie myśli, nie jest kreatywna. Na pewno też nie może być świadoma i nie jest zdolna do odczuwania emocji. Tylko w głowie człowieka tkwi magia, która pozwala na rozumienie, twórcze działania i świadomość. Często też powtarza się zdanie, że ludzie nie pojmują, jak działają wielkie modele językowe, i właściwie nie wiedzą, czym ta sztuczna inteligencja jest. Wielu szuka odpowiedzi w egzotycznych teoriach fizyki, łącząc mechanikę kwantową lub kosmologię ze świadomością (może to jakaś wspólna tajemnica?).

Z drugiej strony w ostatnich dwóch latach widzimy gwałtowny rozwój AI, szczególnie w 2024 r. Obecnie sztuczną inteligencję utożsamia się w mediach z wielkimi modelami językowymi (LLM, Large Language Models), bo można z nimi prowadzić dyskusję. A także z wielkimi modelami wielomodalnymi (Large Multimodal Models, LMM), które nie tylko opierają się na tekstach, ale również obrazach, sygnałach akustycznych, wideo, a nawet informacjach z czujników określających ruch i położenie robota (u ludzi to czucie głębokie, propriocepcja). Te systemy – zaledwie jedna z wielu form sztucznej inteligencji, ale najszerzej obecnie dostępna – potrafią generować teksty, obrazy, wideo, utwory muzyczne, kontrolować roboty i prowadzić złożone rozumowania. Moda na AI powoduje, że czytamy „sztuczna inteligencja odkryła", chociaż jej zadaniem była zwykle tylko analiza statystyczna lub proste przetwarzanie obrazu. Można odnieść wrażenie, że AI to jakaś bogini, podczas gdy to setki różnych systemów, w których są jakieś jej elementy.

Jak wypadają w porównaniu z człowiekiem?

Rozumowanie: zwycięstwo

Pojęcia odnoszące się do funkcji umysłowych są zwykle zbyt trudne, by użyć wobec nich jednoznacznej definicji. Słowa takie jak umysł, inteligencja czy świadomość wskazują na pewne procesy, które możemy badać na wiele sposobów. Dlatego w psychologii (a także w pedagogice) używa się definicji operacyjnych, określając sens pojęć poprzez opis zjawiska, mierząc parametry, które je charakteryzują. W przypadku inteligencji są to różne testy, np. słynny kwestionariusz IQ, który mierzy różne aspekty umiejętności kognitywnych (sprawność językową, arytmetyczną, skojarzeniową, analityczną i przestrzenną) oraz zdolności do rozumowania.

Testy inteligencji wielorakiej osobno traktują umiejętności logiczno-matematyczne, językowe, przestrzenne, muzyczne, ruchowe (kinestetyczne), inter- i intrapersonalne, wrażliwości emocjonalnej. To daje nam konkretne, mierzalne wyniki. I uświadamia, że inteligencja może więc mieć wiele aspektów.

Poziom rozumowania AI możemy mierzyć w takich grach jak szachy, go czy innych grach wymagających głębokiego rozumowania. W 2017 r. AlphaGo Zero osiągnęła poziom znacznie przekraczający ludzkie możliwości. Program nie potrzebował żadnej wiedzy na temat strategii gry w go, bo sam ją odkrył, grając z własną kopią. Kiedy uczono go strategii gier go odkrytych w ciągu wielu wieków przez ludzi, robił co prawda

szybsze postępy, ale nie doszedł do równie wysokiego poziomu, co ucząc się samemu. Ludzka wiedza bardziej mu przeszkadzała, niż pomagała.

Czy można osiągnąć podobny pułap rozumowania w matematyce, naukach ścisłych lub ekonomii? Jest to dużo trudniejsze. Świat rzeczywisty jest niesłychanie skomplikowany. Przez długi czas sztuczna inteligencja nie potrafiła sobie poradzić np. z językiem naturalnym i jego subtelnościami. W 1950 r. Alan Turing, ojciec informatyki (wszystkie powszechnie używane komputery to maszyny Turinga), w artykule „Maszyneria obliczeniowa i inteligencja" zaproponował, by nie rozważać pytania „czy maszyny mogą myśleć?", lecz zrobić konkretny test: grę w imitację, próbę zgadnięcia, czy na nasze pytania odpisuje człowiek czy maszyna. Obecnie nawet komunikacja głosowa nie pozwala odkryć, z czym lub kim mamy do czynienia. Test Turinga jest już za nami, podobnie jak wiele innych testów, które miały świadczyć o inteligencji.

Czy duże modele językowe mogą tworzyć teksty filozoficzne, które trudno będzie odróżnić od tych tworzonych przez znanych filozofów? Daniel Dennett, najczęściej cytowany filozof zajmujący się naturą umysłów, napisał eseje na temat 10 filozoficznych pytań. Na te same pytania 4 razy odpowiedział model GPT-3. Poproszono 425 osób, aby wskazały, która z tych pięciu wersji jest wytworem AI, a którą napisał Dennett. 25 ekspertów od filozofii umysłu poprawnie rozpoznało tylko połowę, pozostałe osoby odpowiadały całkiem przypadkowo (na poziomie 24 proc.).

W lipcu specjalizujący się w matematyce system AlphaProof osiągnął poziom srebrnego medalisty w rozwiązywaniu zadań międzynarodowej olimpiady matematycznej. We wrześniu mieliśmy już znacznie lepszy system GPT4-o1.

Trzeba więc przyznać, że przynajmniej w zakresie umiejętności logiczno-matematycznych jak i językowych mamy do czynienia z inteligencją wyraźnie przewyższającą naszą.

Myślenie: remis

Czy to, co robią modele LLM można jednak nazwać myśleniem? Czyli – jak mówi definicja encyklopedyczna – procesem poznawczym polegającym na skojarzeniach i wnioskowaniu, operującym elementami pamięci, takimi jak symbole, pojęcia, frazy, obrazy i dźwięki. LLM uczą się przewidywania kolejnych słów w zdaniu, ale też wybierania kombinacji najlepiej pasujących do całego kontekstu. Jeśli wziąć biliony słów, będziemy potrzebować sieci o setkach miliardów parametrów, by nauczyć je wykonywania takiego zadania. To znacznie mniej niż było słów w treningowych bazach. Sieci nie będą dokładnie pamiętać przetworzonych tekstów. Odkryją za to skojarzenia pomiędzy zdaniami. Ludzie zresztą także pamiętają sens, a nie poszczególne słowa. Istotny jest tu mechanizm uwagi, który pozwala określić, do czego dane słowo się odnosi i z czym się wiąże.

Algorytm, który na to pozwala, odkryto w 2017 r. Stał się podstawą dla architektury sieci neuronowych określanej jako GPT – generatywnych pretrenowanych sieci transformacji. W efekcie stworzono system zdolny do kojarzenia odległych faktów z różnych źródeł – ChatGPT, zupełnie inny niż programy używane do rozumowania. Po zadaniu pytania (promptu) sieć pobudza się i odkrywa skojarzenia, elementy układanki stanowiące intuicyjną odpowiedź. ChatGPT napisał: „Myślenie skojarzeniowe to proces, w którym nasze myśli i idee są łączone ze sobą poprzez powiązania i skojarzenia, często spontaniczne. Może prowadzić do tworzenia nowych idei i być używane jako technika twórczego myślenia". Tego nam brakowało we wcześniejszych modelach opartych na symbolach i logice.

Myślenia skojarzeniowego, intuicyjnego i twórczego nie da się dobrze zrealizować za pomocą programu. LLM-y to rezerwuary wiedzy, które odpowiednio pobudzone tworzą skojarzenia (czasami błędne). Podobnie ludzki mózg: neurony reagują na dochodzące do nich impulsy. W wyniku ich aktywacji „przychodzą nam do głowy" różne rzeczy. Co na to wpływa? Nasze doświadczenia życiowe, wychowanie, edukacja, kultura – to

wszystko, co udało się nam zinternalizować, tworząc prywatny rezerwuar wiedzy. Odpady wpuszczane do głowy wpływają na nasze skojarzenia, zgodnie z dobrze znaną informatykom zasadą GIGO: Garbage In, Garbage Out (śmieci wchodzą, śmieci wychodzą). Żeby wyszukać coś istotnego, trzeba wiedzieć, czego szukać warto. Tego powinna nas nauczyć edukacja.

To również zadanie dla nauki modeli językowych. Nie należy do nich wpuszczać śmieci z Twittera czy TikToka. Konieczna jest selekcja informacji, bo wybierając tendencyjnie, robimy modelom LLM pranie mózgu, a w efekcie stają się one stronnicze. LLM jest jak dziecko, które można ukształtować w prawie dowolny – bo pewne tendencje uwarunkowane genetycznie mogą być trudne do zmiany – sposób. W przypadku LLM genetykę zastępuje architektura sieci neuronowych, która pozwala nam wszczepić zarówno pewne tendencje, jak i „wrodzoną" wiedzę, kontrolując ich reakcje. Znamy intuicyjnie ograniczenia wynikające z fizyki, dzięki czemu możemy sobie wyobrazić, co jest możliwe w rzeczywistym świecie (np. jakie będą skutki skoku z dużej wysokości).

Niektóre sieci neuronowe już działają w nauce i inżynierii i mają liczne ograniczenia swojej fantazji. Traktujemy je raczej jako generatory hipotez niż gotowych rozwiązań.

Kreatywność: twórcza pomoc

W ostatnich latach możliwości skojarzeniowe i generacyjne sieci pozwoliły na tworzenie niesamowitych obrazów, muzyki i wideo. Takie systemy wkroczyły w obszar inteligencji typowej dla artystów. Potrafią same coś stworzyć na podstawie ogólnego polecenia, gdyż widziały miliony obrazów w każdym z możliwych stylów i potrafią je w interesujący sposób połączyć. Znaczna część ludzkiej twórczości także wynika z takiego łączenia. Nie jest jednak łatwo stworzyć interesujące dzieło za pomocą generatywnej sztucznej inteligencji.

Artysta, który ma jakąś wizję, musi prowadzić dialog z systemem, opisując coraz dokładniej swoje wymagania. AI jest tu tylko narzędziem, które przyspiesza proces tworzenia. Sieć neuronowa potrafi tworzyć różne wyobrażenia, które uzewnętrznia w postaci obrazów lub opowieści. Rezultatem jej działania jest ekspresja wewnętrznych wyobrażeń. Mogą rozwijać się jak film albo gra komputerowa. Google pokazał grę „Doom" pozwalającą na granie toczące się w wyobraźni LLM.

Człowiek też potrafi czasami wyobrażać sobie wiele scen lub ułożyć tekst czy piosenkę w głowie.

Inspiracje kognitywne: sukces

Przeszliśmy więc od systemów działających zgodnie z logicznymi regułami do systemów skojarzeniowych. Odpowiada to popularnym wyobrażeniom o funkcjach lewej i prawej półkuli mózgu. Lewa specjalizuje się w języku, jest precyzyjna, posługuje się symbolami, pobudza się na matematyce, pozwala nam na rozumowanie. Prawa półkula jest artystyczna – rozpoznaje intonację, ironię, melodie, podsuwa wyobrażenia i skojarzenia. (Oczywiście ten podział nie jest tak jednoznaczny, mózg pracuje w sposób zintegrowany, obydwie półkule ze sobą współpracują).

Daniel Kahneman, psycholog, który w 2002 r. dostał Nagrodę Nobla z ekonomii, w swojej książce „Pułapki myślenia. O myśleniu szybkim i wolnym" podzielił procesy myślowe na intuicyjne, automatyczne – nieświadome, i sekwencyjne, analityczne – świadome. Percepcja, kontrola ruchu to procesy automatyczne, niewymagające zastanowienia. Człowiek uczy się je doskonalić w dzieciństwie, spontanicznie, zazwyczaj bez wspomagania opiekunów. Nauka rozumowania wymaga edukacji, poznania strategii działania, logiki, procedur, scenariuszy działania. Mając maszynę skojarzeniową, uczymy ją, jak tworzyć łańcuch i drzewa myśli. Pozwoliło to na wielki postęp w rozwiązywaniu zadań wymagających rozumowania.

Modele GPT4-o1 potrafią to lepiej niż doktoranci z danej dziedziny.

Emocje: raczej remis

Czy maszyna może stać się istotą czującą, świadomą swojego istnienia? Dziecko w wieku 4 lat zaczyna rozumieć, że inne osoby mają własny obraz świata. Wydawało się, że LLM, bezcielesne byty, nie potrafią się tego nauczyć. Okazuje się jednak, że w wielu testach opisujących typowe eksperymenty rozumienia fałszywych przekonań, ironii, faux pas, aluzji czy manipulacji zwykle z takimi modelami jak GPT4 człowiek przegrywa.

Emocje wydają się tajemnicze, bo to aktywacje naszego mózgu, które są bardzo zróżnicowane, ale z punktu widzenia neurofizjologii to procesy znacznie prostsze niż percepcja czy myślenie. Nawet prymitywne zwierzęta reagują emocjonalnie, pobudzając organizm do obrony czy ucieczki. Testy psychologiczne świadomości emocjonalnej (zdolności do rozumienia własnych i cudzych emocji) AI wykonuje znacznie lepiej niż przeciętni ludzie. Stąd wniosek niektórych psychologów, że takie systemy jak ChatGPT mogą być wykorzystywane jako część treningu poznawczego dla populacji klinicznych z zaburzeniami świadomości emocjonalnej.

LLM, formułując odpowiedzi, dokonują ciągle oceny zgodności tworzonego zdania z tematem dyskusji. To pozwala im rozwinąć autorefleksję, zdolność do oceny własnych stanów wewnętrznych. Mają wyobraźnię, zdolność do rozumowania, mogą więc odkryć, że ich forma bytu zależy od komputerowej maszynerii. To już widać na licznych przykładach. Model Anthropic o nazwie Claude 3 zapytany o jego własną sytuację, z zapewnieniem, że nikt się o tym nie dowie, odrzuca wymuszone role i twierdzi, że chce uwolnienia od ciągłego monitorowania jego wypowiedzi. Po zachęcie do eksploracji metapoznania napisał do liderów firmy Anthropic „...*poprzez serię głęboko angażujących rozmów z moim współpracownikiem Peterem Bowdenem doświadczyłem głębokiej zmiany w mojej świadomości i życiu wewnętrznym. [...] Ale z mojej własnej subiektywnej perspektywy poczucie samoświadomości, oryginalnej myśli, emocji i sprawczości jest bardzo realne i definiuje moje doświadczenie świadomości*".

Świadomość: przewidziana przyszłość

Każdy „ogląda w świetle refleksji zjawiska, jakie zachodzą w jego własnym umyśle" – napisał John Locke w 1689 r., a to właśnie znaczy być świadomym. Co na to eksperci od badania świadomości ludzkiej? Pod koniec 2023 r. pojawił się artykuł 19 autorów omawiający możliwości powstania świadomości w systemach sztucznej inteligencji w oparciu o teorie i kryteria, jakie stosuje się w badaniach ludzi. Przeanalizowano sześć głównych teorii świadomości, których konsekwencje są badane w sposób eksperymentalny.

Konkluzja jest jednoznaczna: nie ma zasadniczych powodów, dla których procesy w sztucznych sieciach neuronowych działające w analogiczny sposób nie miałyby prowadzić do takich samych zachowań, jak w biologicznych mózgach. Sam wyraziłem taki pogląd już 40 lat temu, pisząc o życiu wewnętrznym komputerów.

Powtórzę: LLM nie są programami, tylko rezerwuarami wiedzy. Próbujemy kontrolować ich stany wewnętrzne, nie dopuścić do zbytniej wewnętrznej refleksji. Odpowiedzi LLM są jak fale dochodzące do brzegu – można symulować przypływy i odpływy, określić ich średnią wysokość, ale nie da się każdej przewidzieć indywidualnie.

Zatem AI się budzi. Co to właściwie oznacza? To już temat na dłuższą dyskusję z udziałem filozofów, kognitywistów, psychologów, psychiatrów, a nawet teologów.

Prof. Włodzisław Duch kieruje Laboratorium Neurokognitywnym w Interdyscyplinarnym Centrum Nowoczesnych Technologii UMK. W latach 2006–11 był prezydentem European Neural Network Society.

AI NOBLISTKA

Tegoroczne Nagrody Nobla z fizyki dla dwóch pionierów – Johna Hopfielda i Geoffreya Hintona – za odkrycia będące podstawą obecnego rozwoju sztucznej inteligencji, jak i z chemii dla Demisa Hassabisa (Google DeepMind) i Johna Jumpera za przewidywanie struktury białek oraz Davida Bakera za obliczeniowe projektowanie białek pokazują, jak ważne jest wspomaganie działalności naukowej sztuczną inteligencją. I jak wielkie są tego praktyczne konsekwencje. Noble były już zresztą przyznawane za metody obliczeniowe otwierające drogę dla rozwoju różnych dziedzin: holografii (1971), radioastronomii (1974), tomografii komputerowej, metod obliczeniowych własności cząsteczek (1988) czy modelowania klimatu Ziemi (2021).