

Człowiek o AI (10)

* Center for Mind/Brain Sciences,
Department of Information Engineering
and Computer Science, University of
Trento

Autor artykułu, profesor Jakub Szymanik to kognitywista, czyli naukowiec badający działanie ludzkiego umysłu, procesy myślenia oraz sposób, w jaki przetwarzamy język i informacje. Sama kognitywistyka jest interdyscyplinarną dziedziną łączącą m.in. psychologię, językoznawstwo, informatykę i neuronaukę. Badania Jakuba Szymanika koncentrują się szczególnie na relacji między językiem naturalnym, logiką i mechanizmami poznawczymi człowieka.

Rozumowanie sylogistyczne to wnioskowanie oparte na zdaniach łączących dwa pojęcia za pomocą jednego z czterech kwantyfikatorów:

„wszystkie A są B”,
„żaden A nie jest B”,
„niektórzy A są B”,
„niektórzy A nie są B”.

Mając do dyspozycji całą bazę takich zdań, wyprowadzamy nowe wnioski, łącząc je ze sobą zgodnie z regułami logiki. W efekcie rozumowanie polega na systematycznym odkrywaniu, jakie dodatkowe zależności między pojęciami muszą być prawdziwe, skoro prawdziwe są zdania wyjściowe.

Przykład: Załóżmy, że w naszej bazie wiedzy znajdują się zdania:

„wszystkie ssaki są zwierzętami”,
„żaden gad nie jest ssakiem”,
„niektóre zające są ssakami”.

Z tych informacji możemy wywnioskować m.in., że: „niektóre zające są zwierzętami” (bo jeśli coś jest ssakiem, to jest też zwierzęciem).

Czy AI myśli?

Jakub SZYMANIK*

Trudno dziś mówić o sztucznej inteligencji bez napotkania fali entuzjazmu. W mediach, wystąpieniach publicznych i raportach inwestorskich regularnie pojawia się wizja bliskiej przyszłości, w której AI osiąga poziom ogólnej inteligencji (ang. *Artificial General Intelligence*, AGI) i zaczyna zastępować ludzi nie tylko w pracy fizycznej czy rutynowej, ale również w myśleniu, rozumowaniu, a nawet w prowadzeniu badań naukowych.

Tego rodzaju prognozy nie są marginalne. Przeciwnie, jak zauważa Tomasz Kazana we wstępie do tej serii artykułów (patrz Δ_{26}^3), formułują je osoby o ogromnym wpływie. Elon Musk twierdzi, że AI już wkrótce „przewyższy człowieka praktycznie we wszystkim”, a Ray Kurzweil od lat zapowiada nadejście osobliwości technologicznej. Z drugiej strony istnieje równie silny nurt sceptycyzmu, reprezentowany przez naukowców i filozofów, którzy podkreślają ograniczenia obecnych modeli i sugerują, że ich sukcesy mają charakter wyłącznie powierzchowny.

Jak w takim razie odnaleźć się w tej debacie? Jednym ze sposobów jest próba odpowiedzi na bardziej konkretne pytania, na przykład: czy współczesne systemy AI rzeczywiście *rozumują*?

Logika jako test

Jeśli chcemy ocenić zdolność do myślenia, warto sięgnąć po narzędzia, które od wieków służą do jej badania. Jednym z nich jest logika. To właśnie ona stanowi podstawę matematyki i nauki ogólnie, bo pozwala przechodzić od przesłanek do wniosków w sposób uporządkowany i uzasadniony.

Czy modele AI potrafią nauczyć się logiki?

To pytanie stało się przedmiotem naszych badań prowadzonych na Uniwersytecie Warszawskim. Wspólnie z Manuelem Vargasem Guzmánem i Maciejem Malickim analizowaliśmy, czy współczesne modele neuronowe – w tym duże modele językowe – są w stanie opanować systematyczne rozumowanie logiczne. Jako pole testowe wybraliśmy logikę sylogistyczną, sięgającą jeszcze czasów Arystotelesa. Jest ona stosunkowo prosta, a zarazem wystarczająco bogata, by ujawniać subtelne różnice między prawdziwym rozumieniem a powierzchownym dopasowaniem do wzorców.

Dwa oblicza dedukcji

Jak rozpoznać, czy model AI nauczył się logiki, czy tylko udaje? Kluczowe wydają się dwa pojęcia: kompozycyjność i rekurencyjność.

Kompozycyjność oznacza zdolność do rozumienia struktury argumentu – do rozpoznania, że złożone wnioskowania składają się z prostszych kroków, które można analizować niezależnie. Jeśli ktoś pojmuje, że z „wszystkie A są B” i „wszystkie B są C” oraz „wszystkie C są D” wynika, że „wszystkie A są D”, to powinien umieć również rozpoznać, że z „wszystkie A są B” i „wszystkie B są C” wynika, że „wszystkie A są C”.

Rekurencyjność natomiast to zdolność do budowania coraz dłuższych wniosków poprzez powtarzanie tych samych reguł. System rekurencyjny może więc przejść od prostego wnioskowania do bardziej złożonego, dodając kolejne ogniwa do łańcucha. Czyli, jeśli system rozumie, że z „wszystkie A są B” i „wszystkie B są C” wynika, że „wszystkie A są C” oraz jeśli wyabstrahował znaczenie kwantyfikatora „wszystkie”, to powinien też wiedzieć, że z „wszystkie A są B” i „wszystkie B są C” oraz „wszystkie C są D” wynika, że „wszystkie A są D”.

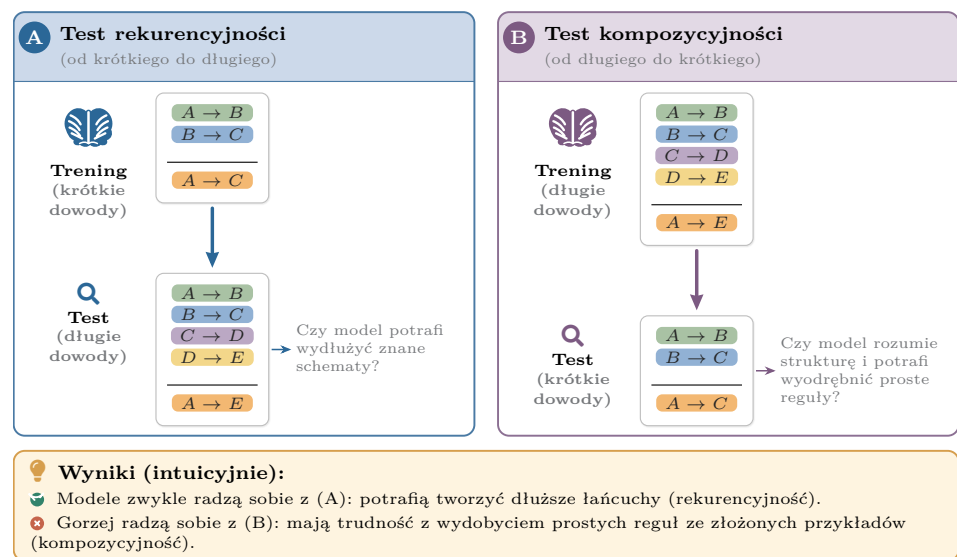
Na pierwszy rzut oka oba te aspekty wydają się niemal tożsame. W praktyce jednak można je rozdzielić. System może być rekurencyjny – generować długie łańcuchy wnioskowań – nie będąc jednocześnie kompozycyjnym, czyli nie rozumiejąc struktury tych wnioskowań.

W jednym z naszych eksperymentów badaliśmy, czy modele AI potrafią uogólniać reguły rozumowania sylogistycznego poza przykłady, na których były trenowane. W tym celu przygotowaliśmy sztuczne „bazy wiedzy” składające się ze zdań sylogistycznych oraz odpowiadających im poprawnych wnioskowań (czyli krótkich lub długich „dowodów”).

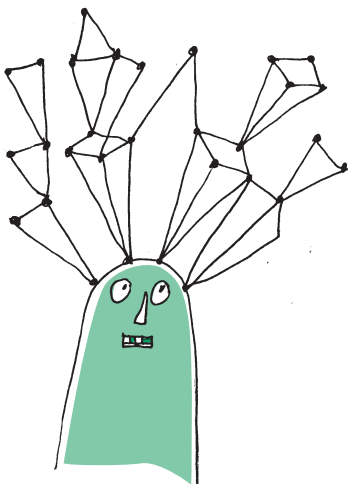
Następnie przeprowadziliśmy dwa kluczowe testy (zob. ilustrację):

- **Test rekurencyjności.** Model trenowaliśmy na krótkich łańcuchach wnioskowań (na przykład złożonych z dwóch–trzech kroków), a następnie sprawdzaliśmy, czy potrafi poprawnie wnioskować w znacznie dłuższych łańcuchach (na przykład pięć–sześć kroków). To pozwalało zbadać, czy umie „powtarzać” znane reguły w coraz dłuższych rozumowaniach.
- **Test kompozycyjności.** Tym razem robiliśmy odwrotnie: model trenowaliśmy na długich, złożonych dowodach, a następnie sprawdzaliśmy, czy potrafi poprawnie rozwiązywać prostsze przypadki. To z kolei testowało, czy model rzeczywiście rozumie strukturę rozumowania i potrafi „rozłożyć” złożony argument na prostsze części.

Eksperyment: rekurencyjność vs. kompozycyjność w rozumowaniu sylogistycznym



MYŚLĘ,
ALE CZY
JESTEM?



Wynik był asymetryczny: modele często radziły sobie całkiem dobrze z pierwszym zadaniem (rekurencyjnością), ale wyraźnie gorzej z drugim (kompozycyjnością). Sugeruje to, że potrafią one naśladować długie ciągi wnioskowań, ale mają trudność z uchwyceniem ogólnych reguł, które za nimi stoją. Co ważne, sprawdzaliśmy zarówno własne modele uczone od podstaw, jak i gotowe, duże modele językowe. Okazało się, że problemy z rozumowaniem występują w obu przypadkach.

Gdzie kończy się „zrozumienie”?

Nasze eksperymenty pokazały, że współczesne modele AI – nawet te najbardziej zaawansowane – mają szczególne trudności z kompozycyjnością. Wprawdzie często potrafią generować poprawnie wyglądające ciągi wnioskowań, ale okazuje się, że robią to raczej na zasadzie statystycznego dopasowania wzorców, a nie poprzez zastosowanie ogólnych reguł.

Innymi słowy: modele potrafią „udawać” rozumowanie, ale nie zawsze je realizują.

To spostrzeżenie wpisuje się w szerszą krytykę głębokich sieci neuronowych oraz dużych modeli językowych, zgodnie z którą działają one na poziomie tekstu, a nie rzeczywistego znaczenia. Jak zauważa Jerzy Marcinkowski w niniejszej serii (Δ_{26}^3), model „zna tylko zdania, nie zna świata”. Jeśli tak, to trudno mówić o klasycznym rozumieniu, które w nauce i filozofii zwykle wiąże się

z odniesieniem do rzeczywistości. W naszym kontekście powiedzielibyśmy, że model nie uczy się na przykład pełnego znaczenia wyrażenia „wszystkie” albo, innymi słowy, nie internalizuje całej „logiki tego wyrażenia”. Modele nie są w stanie wyabstrahować ogólnych reguł przekształcania zdań, w których kwestia odniesienia do świata nie gra roli. To tak, jakby potrafiły mnożyć wielocyfrowe liczby, ale jednocześnie miały problemy z tabliczką mnożenia.

A co z ludźmi?

Ktoś mógłby w tym miejscu zauważyć: przecież ludzie też nie są idealnymi logikami. Istnieje cała dziedzina – psychologia rozumowania – która bada, jak bardzo nasze faktyczne wnioskowania odbiegają od ideału logiki formalnej.

To prawda. Ludzie popełniają błędy, stosują heurystyki, ulegają złudzeniom poznawczym. A jednak mamy do czynienia z istotną różnicą. Nasze rozumowanie nie ogranicza się do operowania symbolami – jest zakorzenione w doświadczeniu świata, percepcji i działaniu. Tymczasem współczesna AI jest w ogromnej mierze „zamknięta” w języku.

Inteligencja fragmentaryczna

To prowadzi do kolejnego pytania: czy myślenie i język są ze sobą nierozzerwalnie związane?

Nie jest to oczywiste. W naukach o poznaniu od lat toczy się debata na ten temat. Tradycja zapoczątkowana przez Noama Chomsky’ego zakładała istnienie wrodzonej struktury językowej, ściśle

powiązanej z naszymi zdolnościami poznawczymi. Z drugiej strony rośnie liczba badań sugerujących, że myślenie może w dużej mierze zachodzić niezależnie od języka – na przykład w formie reprezentacji przestrzennych czy wizualnych, a obwody neuronowe w mózgu odpowiedzialne za język różnią się od tych odpowiedzialnych za różne aspekty myślenia, jak na przykład za logiczne rozumowanie czy teorię umysłu (czyli rozumowanie o przekonaniach, wiedzy lub intencjach innych ludzi).

Jeśli tak jest, to modele językowe – choć imponujące – mogą uchwycić tylko fragment tego, co nazywamy inteligencją.

Wystarczy pomyśleć o zadaniach polegających na manipulowaniu obiektami w przestrzeni, na przykład takich jak testy rotacji brył. Tego rodzaju zdolności trudno sprowadzić do operacji na tekście. A jednak stanowią istotny element ludzkiego poznania.

AI jako eksperyment myślowy

Jak zatem interpretować sukcesy sztucznej inteligencji?

Być może najrozsądniejsze podejście polega na traktowaniu jej jako swoistego eksperymentu. Modele AI nie są prostym odwzorowaniem ludzkiego umysłu, lecz jedną z wielu możliwych realizacji procesów poznawczych. Pokazują, że pewne funkcje – takie jak generowanie języka czy rozwiązywanie zadań – mogą być osiągnięte poprzez analizę statystycznych zależności w danych. To samo w sobie jest fascynujące, bo poszerza naszą wyobraźnię o tym, czym może być inteligencja.

Więcej o badaniach prof. Jakuba Szymanika można przeczytać w jego pracach:

- Manuel Vargas Guzmán, Jakub Szymanik, Maciej Malicki. *Testing the limits of logical reasoning in neural and hybrid models*. Findings of the Association for Computational Linguistics: NAACL, str. 2267–2279. Association for Computational Linguistics, 2024.
- Leonardo Bertolazzi, Manuel Vargas Guzmán, Raffaella Bernardi, Maciej Malicki, Jakub Szymanik. *Teaching small language models to learn logic through meta-learning*. Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers), pages 8049–8080. Association for Computational Linguistics, 2026.

Jednocześnie jednak nie oznacza to, że jesteśmy o krok od stworzenia systemów, które „myślą” w takim sensie jak ludzie. Jak zauważa w tej serii Piotr Sankowski (Δ_{26}^3), nawet najbardziej zaawansowane systemy wymagają ludzkiej interpretacji i współpracy, a ich rola polega raczej na wspieraniu procesu badawczego niż na jego zastępowaniu.

Między szumem medialnym a nauką

Debata o AI toczy się dziś na wielu poziomach – naukowym, filozoficznym, ekonomicznym, a nawet geopolitycznym. Nie można przy tym zapominać, że ogromna część rozwoju tej technologii odbywa się w sektorze prywatnym, gdzie naturalną rolę odgrywają marketing i budowanie narracji. Innymi słowy, istotna część szumu medialnego służy po prostu nakręcaniu zysków inwestorów i właścicieli firm.

To nie znaczy, że postęp jest iluzją. Przeciwnie – jest realny i imponujący. Warto jednak zachować ostrożność w interpretowaniu jego znaczenia. Zamiast pytać, czy AI „już myśli”, być może lepiej zapytać: *jakie aspekty myślenia udało się uchwycić, a jakie wciąż pozostają poza zasięgiem?*

Zamiast zakończenia

Sztuczna inteligencja nie jest ani magicznym bytem, który zaraz nas zastąpi, ani zwykłym narzędziem bez głębszego znaczenia. Jest raczej lustrem, w którym odbijają się nasze intuicje dotyczące inteligencji, języka i poznania.

Patrząc w to lustro, uczymy się nie tylko o maszynach, ale również o nas samych.

I być może najciekawsze nie jest dziś to, czy AI myśli, lecz to, co jej sukcesy mówią nam o naturze ludzkiego myślenia.