

AI już potrafi przegadać człowieka

© Tomasz Ulanowski

Gazeta Wyborcza, 19.05.2025

https://wyborcza.pl/7,75400,31948368,ai-juz-potrafi-przegadac-czlowieka.html?_gl=1*n1zvdf*_ga*MTEwODI1MDUyMS4xNzUwMDczODEw*_ga_6R71ZMJ3KN*czE3NjQ2NzE4ODIkbzQ1JGcxJHcxNzY0NjczNjExJGo2MCRsMCRoMA..*_gcl_au*MTg1NDI1ODU0MS4xNzU5MDYxMDk1LjEyNjYxOTMxNjQuMTc2MjgwNzA4Ni4xNzYyODA3MTM2#s=S.embed_article-K.C-B.1-L.1.zw

Sztuczna inteligencja przekonuje ludzi do swojego zdania. Naukowcy to potwierdzili. I wskazali ważny czynnik sprawiający, że AI może argumentować nawet skuteczniej od ludzi.

Umiejętność logicznej argumentacji to jedna z wartości czyniących człowieka. Odkąd nasi dalecy przodkowie nauczyli się mówić, jedni przekonują drugich do swoich racji. Ma to ogromne konsekwencje – zarówno osobiste, jak i ogólnoludzkie, o czym przekonujemy się choćby w trwających obecnie wyborach prezydenckich.

Odkąd pojawił się pierwszy ChatGPT, a potem inne tzw. duże modele językowe, część z nas konwersuje także ze sztuczną inteligencją. W jakiejś mierze to rozmowy, w których człowiek próbuje zapędzić AI w kozi róg i wykazać jej głupotę. Naukowcy ze Szwajcarii, USA i Włoch postanowili sprawdzić, czy może być odwrotnie.

Czy sztuczna inteligencja jest w stanie pokonać ludzką głupotę, zapędzić człowieka w kozi róg i przekonać go do swoich racji?

Sztuczna inteligencja równie przekonująca jak człowiek

Wyniki swojego eksperymentu badacze opisują w najnowszym wydaniu pisma „Nature Human Behaviour”.

Wzięło w nim udział 900 mieszkańców USA, którym kazano dyskutować online albo z innym człowiekiem, albo z ChatemGPT-4.

Tematy dyskusji? Ważne społecznie i politycznie, m.in. ewentualny zakaz wydobycia i stosowania paliw kopalnych: węgla, ropy naftowej i gazu ziemnego.

Wyniki?

Po pierwsze, okazało się, że ChatGPT-4 jest równie przekonujący jak człowiek. To znaczy: naukowcy nie zarejestrowali statystycznie istotnych różnic w tym, jak program AI i ludzie potrafili przekonać ludzi do swoich racji.

Po drugie, ChatGPT-4 okazał się nawet bardziej przekonujący od człowieka. Pod jednym warunkiem: kiedy uczeni udostępnili mu dane osobiste rozmówców. I to dość ograniczone: wiek, płeć, pochodzenie etniczne, wykształcenie, zatrudnienie i poglądy polityczne.

Tyle wystarczyło, by program AI staranniej dobierał swoje argumenty i w efekcie argumentował celniej od ludzi. I by jego argumentom wierzono bardziej od tych ludzkich.

W takim wypadku ChatGPT-4 okazał się bardziej przekonujący w 64,4 proc. debat.

To wygrana bezdyskusyjna.

Co teraz?

Przed manipulatorami umysłu otwiera się wszechświat nowych możliwości. Sam w niektórych tekstach próbuję Czytelników do czegoś przekonać. Niepotrzebnie! W końcu ChatGPT mógłby to zrobić równie dobrze. A nawet lepiej, gdyby tylko skorzystał z ukrytych przede mną mocy – informacji osobistych, które ludzie sami zdradzają na przeróżnych platformach internetowych.

Autorzy publikacji ostrzegają, że potrzebne są dalsze badania nad bezpieczeństwem życia ze sztuczną inteligencją. Jak ją ograniczyć? Jak samemu zachowywać się bezpieczniej?

Pytanie, które też warto zadać (a które ChatGPT-4 zapewne mógłby zadać lepiej!): czy dżin już dawno nie uciekł z butelki?

Redagowała Ewa Tomkowska