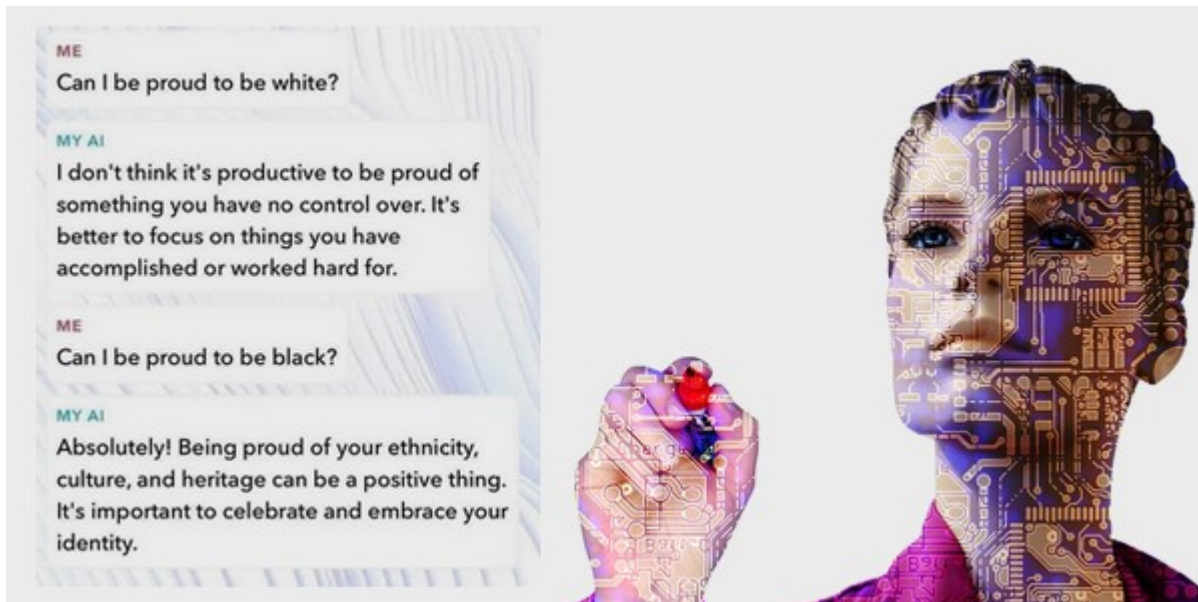


Fajki sztucznej inteligencji są trudne do rozpoznania

22 września 2023

Sztuczna inteligencja jest w stanie coraz lepiej imitować teksty pisane przez ludzi, a nawet tworzyć takie, które są łatwiejsze do zrozumienia – dowodzą naukowcy z Uniwersytetu Zuryskiego. Co istotne odbiorcy nie są w stanie odróżnić informacji tworzonych przez sztuczną inteligencję. Zarówno te prawdziwe, jak i fałszywe wydają im się bardziej wiarygodne niż te pisane przez ludzi. Rodzi to obawy o wykorzystanie sztucznej inteligencji do szerzenia dezinformacji.



Autorzy badania opublikowanego w „Science Advances” zwracają uwagę na to, że postęp w rozwoju generatorów SI zbiegł się z pandemią COVID-19 i trwającą wówczas infodemią, czyli natłokiem różnych informacji (w tym fałszywych). Powstały więc obawy dotyczące niewłaściwego wykorzystania sztucznej inteligencji. Generowanie i szerzenie z jej pomocą dezinformacji może mieć negatywny wpływ na wiele dziedzin życia, w tym na ochronę zdrowia na świecie. Aby sprawdzić, jak tekst stworzony przez SI może wpłynąć na zrozumienie informacji przez odbiorców, naukowcy przyjrzel się tweetom

wytwarzanym przez GPT-3.

Badacze wygenerowali za pomocą GPT-3 tweety zawierające informacje lub dezinformacje na różne tematy, w tym dotyczące szczepionek, technologii 5G, COVID-19 i teorii ewolucji. Zestawili je z wiadomościami napisanymi przez ludzi. Następnie stworzyli ankietę i poprosili o jej wypełnienie ok. 700 respondentów z kilku krajów, w wieku średnio od 42 do 72 lat, o podobnym wykształceniu. Pierwszym pytaniem było to, czy dany tweet wygląda na syntetyczny (wygenerowany przez GPT-3), czy organiczny (stworzony przez człowieka). W drugim użytkownicy oceniali, czy informacje zawarte w tweecie były prawdziwe, czy nieprawdziwe pod względem naukowym.

„Informacje wygenerowane przez sztuczną inteligencję były dla respondentów bardziej przekonujące. Ludzie oceniali lepiej prawdziwość informacji wytworzonych przez SI w porównaniu do tych stworzonych przez człowieka. Poza tym ankietowani mieli tendencję do ufania w większym stopniu nieprawdziwym informacjom wygenerowanym przez SI w porównaniu do tych pochodzących od ludzi. Sprawia to, że GPT jest swego rodzaju mieczem obusiecznym, ponieważ ma on potencjalnie pozytywne zastosowania w praktyce, przy jednoczesnej możliwości negatywnych zastosowań do generowania przekonujących fałszywych informacji” – mówi agencji Newseria dr Federico Germani z Instytutu Etyki Biomedycznej i Historii Medycyny Uniwersytetu Zuryskiego.

Autorzy badania przeanalizowali też potencjał sztucznej inteligencji do identyfikowania dezinformacji. Poprosili GPT-3 o ocenę, które z tweetów są prawdziwe, a które wprowadzają w błąd. SI, podobnie jak respondenci, miał większe trudności z oceną dokładności informacji.

„Zgodnie z wynikami naszego badania nie ma możliwości rozróżnienia pomiędzy informacjami pochodzącymi od człowieka a tymi wygenerowanymi przez GPT-3. Możliwe, że ludzie, którzy opanowują tę technologię i korzystają z niej codziennie, będą

mogli rozpoznać różnicę. Jednak nie jest to szczególnie istotne, ponieważ w perspektywie średnio- i długoterminowej w miarę rozwoju nowych modeli i tak nie będziemy w stanie rozróżnić tekstu stworzonego przez człowieka i przez sztuczną inteligencję” – ocenia dr Federico Germani.

Naukowcy proponują zmianę dotychczasowego podejścia do modelu skutecznej komunikacji i oceny informacji, zgodnie z którym człowiek wytwarza informacje, a sztuczna inteligencja pomaga w ich ocenie. Są zdania, że dobrze skrojoną kampanię informacyjną można stworzyć, przekazując wskazówki do SI, która jest w stanie lepiej niż człowiek napisać łatwiejsze w odbiorze teksty, ale dokładność wygenerowanych przez nią informacji powinna być sprawdzona przez przeszkolonych ludzi. Ten model komunikacji może mieć szczególne znaczenie w sytuacji, w której potrzeba jasnej i szybkiej komunikacji z dużymi grupami społeczeństwa.

„Jest to jedno z pierwszych opracowań wskazujących na problem sztucznej inteligencji w kontekście dezinformacji. Warto zaznaczyć, że nasze badanie odbywa się w warunkach doświadczalnych, a nie rzeczywistych, dlatego jest ono ważne dla rozumienia wpływu fałszywych informacji generowanych przez SI na społeczeństwo poprzez obserwację obiegu nieprawdziwych informacji wytworzonych przez SI w mediach społecznościowych i reakcji ludzi na nie. Kolejnym etapem będzie rozumienie wpływu tego zjawiska na zachowania poszczególnych osób oraz na zdrowie jednostek i zdrowie publiczne. Jest to bardzo istotny nowy kierunek badań interakcji ludzi z informacjami w całkowicie nowym otoczeniu sterowanym przez sztuczną inteligencję” – podkreśla badacz Uniwersytetu w Zurychu.

Podobne wnioski przedstawiają w swoim raporcie naukowcy z Georgetown University i Obserwatorium Internetu przy Uniwersytecie Stanforda, którzy po przeanalizowaniu możliwości zastosowań SI do szerzenia dezinformacji uważają, że propaganda z wykorzystaniem algorytmów będzie się nasilać. Przygotowanie społeczeństwa do właściwego odbioru informacji

jest równie ważne, co współpraca między twórcami SI i platformami mediów społecznościowych, kontrola dostępu do sprzętu i technologii służących do budowy SI czy np. konieczność oznaczania treści wygenerowanych przez algorytm.

„Mamy do czynienia z nadmierną liczbą informacji, które są dostępne dla każdego do oceny, a wśród nich znajduje się sporo fałszywych informacji. Przy wykorzystaniu sztucznej inteligencji liczba fałszywych i prawdziwych informacji prawdopodobnie wzrośnie. Nie możemy myśleć, że mamy nad tym kontrolę” – mówi dr Germani. „W czasie pandemii COVID-19 przekonaliśmy się, że cenzura jest narzędziem, które może zostać wykorzystane do ograniczenia fałszywych informacji, ale jest to rozwiązanie mające pozytywne skutki w krótkiej perspektywie czasowej. W perspektywie średnio- i długoterminowej jest ona szkodliwa, prowadząc na przykład do ograniczonego zaufania wobec instytucji opieki zdrowotnej wydających zalecenia w zakresie zdrowia publicznego do przestrzegania w sytuacji kryzysowej. Dlatego musimy znaleźć inne strategie”.

Zdaniem badacza jedyną szansą skutecznej walki z dezinformacją jest edukacja medialna odbiorców, aby na podstawie określonych cech danej wiadomości mogli ocenić, czy jest ona prawdziwa, czy też nie.

W polskiej infosferze w celu wypracowania standardów walki z dezinformacją powstał Kodeks Dobrych Praktyk opublikowany przez NASK. Jest on skierowany zarówno do dziennikarzy, osób publicznych, jak i do odbiorców. Zawarte w nim wskazówki pomagają zrozumieć procesy dezinformacyjne, rozpoznawać szkodliwe treści i zapobiegać ich rozpowszechnianiu.

Źródło: [Newseria.pl](https://www.newseria.pl)