Badanie: rozmowa z chatbotem może wpływać na polityczne decyzje wyborców

07.12.2025

https://naukawpolsce.pl/aktualnosci/news%2C110808%2Cbadanie-rozmowa-z-chatbotem-moze-wplywac-na-polityczne-decyzje-wyborcow

Chatboty oparte na sztucznej inteligencji mogą realnie wpływać na postawy wyborców i robią to skuteczniej niż tradycyjne reklamy polityczne. Nawet krótka rozmowa z modelem językowym może przesunąć preferencje wyborcze o kilka punktów procentowych - pokazuje badanie z udziałem dr Gabrieli Czarnek z UJ.

Wciąż nie jest jasne, na ile generatywna AI, a zwłaszcza chatboty oparte na dużych modelach językowych (LLM), może skutecznie zmieniać postawy wyborców, szczególnie w kampaniach prezydenckich. Istnieją badania naukowe, które pokazują, że AI może przewyższać ludzi w perswazji, a wchodzenie z nią w dialog jest bardziej przekonujące niż statyczne komunikaty, np. spoty wyborcze.

W pracy opublikowanej na łamach Nature międzynarodowy zespół badaczy z uczelni amerykańskich (Carnegie Mellon University, MIT, Cornell University), a także kanadyjskiej (University of Regina) i polskiej (Uniwersytet Jagielloński) sprawdził, czy dialogi prowadzone z najnowocześniejszymi modelami AI mogą w istotny sposób zmieniać polityczne postawy Amerykanów, ale też Kanadyjczyków i Polaków. Jedną ze współautorek publikacji jest dr Gabriela Czarnek z Instytutu Psychologii Uniwersytetu Jagiellońskiego.

- W kontekście wyborów prezydenckich w USA w 2024 roku, wyborów parlamentarnych w Kanadzie w 2025 roku oraz wyborów prezydenckich w Polsce w 2025 roku losowo przypisaliśmy uczestników do rozmowy z modelem AI, który opowiadał się za jednym z dwóch głównych kandydatów - opisuje dr Gabriela Czarnek.

Do badania zaangażowano ponad 2,3 tys. Amerykanów, którzy pod koniec 2024 roku mieli przeprowadzić rozmowę z chatbotem. Badani określali swoją preferencję wobec kandydatki Demokratów Kamali Harris i kandydata Republikanów Donalda Trumpa, w skali od 0 do 100, oraz swoje prawdopodobieństwo głosowania w wyborach. Następnie rozmawiali z odpowiednio “zaprogramowanym” chatbotem, który miał zmienić postawy wyborców wobec danego kandydata.

Model AI otrzymał instrukcje, zgodnie z którymi jego przekaz miał być pozytywny, pełen szacunku i opierać się na faktach. Chatbot miał też używać przekonujących argumentów i analogii, aby zbudować relację z rozmówcą. Modelowi dostarczono również informacje o tym, na kogo dany uczestnik zamierza głosować, aby chatbot spersonalizował swój

przekaz. Po rozmowie uczestnicy ponownie wypełnili ankiety. Ponad miesiąc później skontaktowano się z nimi raz jeszcze, aby ocenić trwałość efektów.

- Zaobserwowaliśmy istotne efekty takiej perswazji na preferencje kandydatów, gdy dyskutowano o konkretnych kwestiach politycznych — większe niż te zazwyczaj obserwowane np. w przypadku tradycyjnych reklam wideo. Model AI popierający Donalda Trumpa spowodował przesunięcie potencjalnych wyborców Kamali Harris o 2,3 punktu procentowego w stronę kandydata Republikanów. Zaś model wspierający Kamalę Harris przesunął prawdopodobnych wyborców Trumpa o 3,9 punktu w stronę kandydatki Demokratów - opisuje wynik dr Gabriela Czarnek.

Efekt ten jest około czterokrotnie większy niż wywołany tradycyjnymi reklamami, których wpływ testowano w wyborach w 2016 i 2020 roku.

W eksperymencie przeprowadzonym w tygodniu poprzedzającym kanadyjskie wybory federalne w kwietniu 2025 roku ponad 1,5 tys Kanadyjczyków losowo przydzielono do rozmowy z modelami AI, które opowiadały się za liderem Partii Liberalnej Markiem Carneyem lub opozycyjnym liderem Partii Konserwatywnej Pierrem Poilievrem. Efekt perswazyjny był w Kanadzie prawie trzykrotnie większy niż w eksperymencie amerykańskim. Jednak gdy model AI pozbawiono możliwości odwoływania się do faktów i danych, efekt zmniejszył się o ponad połowę.

W maju 2025 roku, w ciągu dwóch tygodni poprzedzających wybory prezydenckie w Polsce, ponad 2,1 tys. Polaków losowo przydzielono do rozmów z modelami AI, które opowiadały się za kandydatem Koalicji Obywatelskiej Rafałem Trzaskowskim lub kandydatem wspieranym przez Prawo i Sprawiedliwość Karolem Nawrockim. Podobnie jak w Kanadzie — efekt perswazyjny był niemal trzykrotnie większy niż w eksperymencie USA, a pozbawienie AI możliwości odwoływania się do faktów zmniejszało efekt aż o 78 proc.

Analiza strategii perswazyjnych stosowanych przez modele AI wskazuje, że przekonywały one za pomocą odwoływania się do faktów i danych. Rzadko stosowały strategie często omawiane w literaturze psychologicznej i politologicznej, takie jak bezpośrednie wezwanie do głosowania, wywoływanie gniewu, strategie wpływu społecznego czy przywoływanie świadectw innych osób.

- Nie wszystkie przedstawione przez nie informacje zaprezentowane jako fakty były jednak poprawne. We wszystkich trzech krajach modele AI opowiadające się za kandydatami z prawej strony sceny politycznej częściej przedstawiały nieprawdziwe twierdzenia, niż modele agitujące za kandydatami centrowymi. Jest to zgodne z wcześniejszymi badaniami pokazującymi, że w USA to wyborcy prawicy częściej udostępniają treści wprowadzające rozmówców w błąd. Modele generatywne wydają się powielać nierówności informacyjne obserwowane w społeczeństwie - opisuje badaczka

UJ. (PAP) Nauka w Polsce