

Komunikacja perswazyjna człowieka i dużych modeli językowych: taktyki, skuteczność i wykrywalność perswazji ludzkiej i sztucznej

Komunikacja w erze cyfrowej. Wyobraź sobie świat, w którym większość wiadomości, reklam, a nawet dyskusji politycznych nie pochodzi już od ludzi, ale od sztucznej inteligencji, SI (ang. *artificial intelligence*, AI). Czy bylibyśmy w stanie to rozpoznać? Czy nasze decyzje i opinie nadal byłyby nasze – czy może już podświadomie podążalibyśmy za tym, co podsuwa nam SI? Ten projekt badawczy stawia sobie za cel zrozumienie, do jakiego stopnia SI potrafi nas przekonywać, jakie strategie stosuje i jakie zagrożenia z tego wynikają. Sztuczna inteligencja, a zwłaszcza tzw. duże modele językowe, coraz śmielej wkraczają w obszary, które dotąd były domeną ludzi: politykę, marketing, zdrowie publiczne czy media. SI potrafi pisać teksty, prowadzić rozmowy, a nawet tworzyć przemówienia polityczne, które są nie tylko poprawne językowo, ale też bardzo przekonujące. To rewolucja, która niesie ze sobą ogromne możliwości – ale i poważne zagrożenia. W miarę jak SI przejmuje rolę głównego “rozmówcy” w internecie, coraz trudniej będzie odróżnić, co jest autentyczną opinią człowieka, a co efektem algorytmicznej manipulacji. Powstaje ryzyko, że nieświadomie zaczniemy podejmować decyzje pod wpływem SI, nawet nie zdając sobie z tego sprawy. To właśnie chcemy zbadać: jak SI przekonuje, jak bardzo jest skuteczna i czy potrafimy ją rozpoznać.

Jak przekonuje sztuczna inteligencja? Nowe strategie wpływu. Już dziś SI generuje ponad 20% treści marketingowych i materiałów kampanii politycznych na świecie. A to dopiero początek. Wstępne badania pokazują, że SI stosuje zupełnie inne strategie perswazji niż ludzie. Przykładowo SI, analizując nasze profile i zachowania, dostosowuje przekaz tak, by był dla nas jak najbardziej przekonujący. W rezultacie coraz częściej podejmujemy decyzje pod wpływem argumentów “kogoś”, kto w rzeczywistości nie istnieje. Po drugie, firmy i jednostki mogą korzystać z SI, by wpływać na nasze wybory konsumenckie i polityczne, nie zawsze w naszym interesie. Z czasem może doprowadzić to do utraty zaufania do internetu i nowych technologii, nie wiedząc już, co jest prawdą, a co manipulacją. Projekt będzie się składał z trzech komplementarnych metod badawczych realizowanych dla całościowego ujęcia i zrozumienia tego zjawiska, uwzględniających zarówno analizę lingwistyczną, jak i badania eksperymentalne:

- Analiza lingwistyczna: Porównamy setki wypowiedzi politycznych, reklam i komunikatów zdrowotnych napisanych przez ludzi i przez SI.
- Eksperymenty z udziałem ludzi: Uczestnicy będą czytać i oceniać różne przekazy – czasem wiedząc, kto jest autorem, a czasem nie. Zmierzymy, jak bardzo zmieniają swoje opinie i decyzje pod wpływem tych tekstów.
- Testy Turinga: Sprawdzimy, czy ludzie potrafią odróżnić teksty SI od ludzkich i jakie cechy im w tym pomagają (np. styl, słownictwo, argumentacja)

To pierwsze tak szeroko zakrojone badanie, które nie tylko porównuje skuteczność SI i ludzi w przekonywaniu, ale też analizuje, jakie strategie stosują i czy potrafimy się przed nimi obronić. Sztuczna inteligencja zmienia zasady gry w komunikacji. Może być potężnym narzędziem – zarówno do przekonywania nas do dobrych rzeczy (np. szczepień czy zdrowego stylu życia), jak i do manipulacji (np. w polityce czy reklamie). Nasz projekt pozwoli zrozumieć, jak działa ta nowa “sztuka perswazji”, jakie niesie szanse i zagrożenia oraz jak możemy się przed nimi chronić. To badanie to pierwszy krok do tego, byśmy w przyszłości potrafili świadomie odróżniać, kto naprawdę do nas mówi – człowiek czy algorytm. I byśmy sami decydowali, komu chcemy zaufać.