

# Sztuczna inteligencja z uprzedzeniami

31 lipca 2021

Zespół badaczy z University of Toronto bada występowanie stereotypów i uprzedzeń w systemach opartych na sztucznej inteligencji (Artificial Intelligence, AI). Naukowcy oferują bezpłatną usługę sprawdzenia, czy dany system „jest uprzedzony” i w jakim stopniu.



Profesor Parham Aarabi prowadzący projekt HALT mówi, że uprzedzenia i jakaś forma dyskryminacji występują w niemal każdym testowanym systemie. Podkreśla, że takie systemy mogą być niesprawiedliwe i nawet mogą wpływać na czyjeś życie. Jego grupa zwraca szczególną uwagę na uwzględnienie i rozpoznawanie różnorodności społecznej.

Systemy oparte na sztucznej inteligencji używane są w wielu miejscach, i to coraz powszechniej. Działają np. na lotniskach, w instytucjach rządowych, w agencjach zdrowia, w policji, w telefonach komórkowych, mediach społecznościowych, a czasem nawet w czasie procesów rekrutacyjnych. Wiadomo jednak, że projektują je ludzie – i to jest źródłem uprzedzeń systemów.

Jednym z etapów tworzenia systemu opartego na sztucznej inteligencji jest proces „uczenia się”. Wtedy wykorzystywane są np. zdjęcia lub próbki dźwięku, by w systemie powstały odpowiednie powiązania i skojarzenia. Jeśli system ma rozpoznawać twarze, a w czasie nauki będzie analizował wizerunki osób np. tylko z jednej grupy etnicznej, to wobec innych będzie „uprzedzony” – czyli będzie pokazywał błędy. Grupa profesora Aarabiego zajmowała się też systemami rozpoznawania mowy. Tu było widać, że niektóre akcenty czy dialekty są rozpoznawane dużo gorzej niż mowa osoby, dla której angielski jest językiem ojczystym. Dla osób mówiących z akcentem system stawał się mało użyteczny.

Pełny proces oceny wykonany przez zespół HALT może trwać nawet dwa tygodnie. Po takiej ewaluacji właściciel systemu wie, ile uprzedzeń siedzi w jego technologii. Naukowcy potrafią dokładnie wskazać, jaka grupa społeczna została pominięta albo w inny sposób jest poszkodowana. Na tej podstawie można następnie dobrać dane do „nauki” systemu, by uczynić go sprawiedliwszym.

Autorstwo: Katarzyna Nowosielska

Ilustracja: [Geralt \(CC0\)](#)

Źródło: [Goniec.net](#)