
Sztuczna inteligencja nie może powtarzać hejtu jako faktów

Oficjalne stanowisko Fundacji Usuwanie Hejtu. Stanowisko Fundacji Usuwanie Hejtu: odpowiedzi generowane o ludziach i firmach powinny wskazywać źródła i dawać prostą drogę do sprostowania.

Coraz więcej osób pyta o firmy i o ludzi czaty i asystentów sztucznej inteligencji, a wyszukiwarki coraz częściej same podają gotową odpowiedź. Te odpowiedzi nie biorą się znikąd. Powstają z treści, które już są w sieci: z artykułów, opinii, wpisów na forach i w mediach społecznościowych. Jeśli w sieci jest kłamstwo o firmie albo o człowieku, odpowiedź może je powtórzyć spokojnym, pewnym tonem, jakby było faktem.

Dla osoby zaatakowanej hejtem to nowy wymiar szkody. Do tej pory szkodliwy wpis trzeba było znaleźć. Dziś wystarczy zapytać, a maszyna poda go w streszczeniu, bez kontekstu, bez informacji, że autor był anonimowy, a wpis został zgłoszony albo usunięty. Firma albo człowiek nie ma z kim rozmawiać, bo odpowiedź nie ma autora.

Fundacja Usuwanie Hejtu stoi na stanowisku, że odpowiedzi generowane o konkretnych ludziach i firmach powinny spełniać trzy warunki. Po pierwsze: wskazywać źródła, z których pochodzą, tak żeby użytkownik mógł sprawdzić, skąd wzięła się dana informacja. Po drugie: odróżniać fakty od opinii i wyraźnie zaznaczać, gdy informacja pochodzi z anonimowego albo niezweryfikowanego źródła. Po trzecie: dawać prostą drogę do sprostowania, czyli miejsce, w którym osoba lub firma może zgłosić nieprawdziwą treść i w rozsądnym czasie otrzymać odpowiedź.

Nie chodzi o cenzurowanie krytyki. Rzetelna, sprawdzalna krytyka ma prawo istnieć także w odpowiedziach sztucznej inteligencji. Chodzi o to, żeby narzędzia, z których korzystają miliony ludzi, nie zamieniały nagonki w encyklopedię.

Fundacja przygotowuje Instytut badań nad hejtem i dezinformacją, który ma badać między innymi to, jak odpowiedzi sztucznej inteligencji opisują ludzi i firmy. Jesteśmy gotowi dzielić się wynikami z twórcami tych narzędzi, z mediami i z instytucjami. Osoby i firmy, które już dziś znajdują nieprawdziwe informacje o sobie w takich odpowiedziach, mogą zgłosić się do Fundacji przez stronę usuwaniehejtu.pl.

Fundacja Usuwanie Hejtu · Łukasz Wudyka, prezes · media@usuwaniehejtu.pl · infolinia 222 900 142 · usuwaniehejtu.pl