

Zaapeluj do firm OpenAI, Google i Microsoft, aby zapewniły przejrzystość danych używanych do trenowania sztucznej inteligencji!

Modele generatywnej sztucznej inteligencji są trenowane przy użyciu terabajtów danych pozyskanych z różnych miejsc w internecie. Jednym z najbardziej popularnych źródeł danych treningowych jest mała organizacja non profit o nazwie Common Crawl, która stworzyła ogromne archiwum danych pozyskanych z internetu. W toku najnowszego dochodzenia Mozilla odkryła, że dzięki organizacji Common Crawl opracowywanie generatywnej AI jest bardziej transparentne i konkurencyjne. Jednak twórcy sztucznej inteligencji muszą zapewnić przejrzystość również w zakresie sposobu wykorzystywania takich danych, ponieważ zbiory te odzwierciedlają uprzedzenia użytkowników internetu i zawierają treści, które są toksyczne i szkodliwe. **Tymczasem nie wiemy nawet, czy firmy, takie jak Microsoft, Google i Meta, korzystają z danych Common Crawl do trenowania produktów opartych na AI. Nie mówiąc już o tym, czy szkodliwe treści są odfiltrowywane.**

Lepsze i godne zaufania produkty oparte na sztucznej inteligencji naprawdę są możliwe. Musimy wiedzieć, w jaki sposób sztuczna inteligencja jest trenowana, aby zrozumieć jej ograniczenia i ryzyko związane z korzystaniem z niej. Ale przede wszystkim firmy muszą tworzyć rozwiązania, którym można zaufać i które stanowią pomoc dla wszystkich użytkowników internetu.

Podpisz petycję Mozilli i powiedz firmom OpenAI, Google, Microsoft i Meta, aby zapewniły przejrzystość danych używanych do trenowania narzędzi opartych na sztucznej inteligencji!