

# Zrozumienie wpływu interakcji człowieka ze sztuczną inteligencją na dyskryminację

Posted on 2025-02-28

Ponieważ systemy sztucznej inteligencji są coraz częściej wykorzystywane do wspomagania podejmowania decyzji w sektorach o wysokiej stawce, takich jak kredyty i rekrutacja, eksperci ds. projektowania i analizy behawioralnej z Laboratorium Polityki UE postanowili zbadać kluczową kwestię ludzkiego nadzoru nad decyzjami wspieranymi przez sztuczną inteligencję. Przy wsparciu Europejskiego Centrum Przejrzystości Algorytmicznej (ECAT) Komisja Europejska starała się lepiej zrozumieć, w jaki sposób ludzie wchodzą w interakcje ze sztuczną inteligencją.

## Sprawiedliwość w AI: bardziej złożona niż

# sądziłyśmy

Intuicyjnie można by sądzić, że ludzki nadzór może służyć jako kontrola przed uprzedzeniami AI. Jednak kompleksowe badanie Komisji Europejskiej, które łączy w sobie badania ilościowe i jakościowe, opowiada inną historię. KE odkryła, że nadzorujący pracę ludzi są tak samo skłonni postępować zgodnie z radami systemów AI, niezależnie od tego, czy są one zaprogramowane pod kątem uczciwości, czy nie. Sugeruje to, że sam ludzki nadzór jest niewystarczający, aby zapobiec dyskryminacji; w rzeczywistości może nawet ją utrwalać.

W opisywanym eksperymencie ilościowym wzięli udział specjaliści ds. HR i bankowości z Włoch i Niemiec, którzy podejmowali decyzje dotyczące zatrudniania i udzielania kredytów pod wpływem zaleceń sztucznej inteligencji. Wyniki były dość uderzające: użycie „sprawiedliwej” sztucznej inteligencji zmniejszyło na przykład uprzedzenia ze względu na płeć, ale nie wyeliminowało wpływu wcześniej istniejących ludzkich uprzedzeń w podejmowaniu decyzji. Jakościowe spostrzeżenia z wywiadów i warsztatów z uczestnikami i ekspertami ds. sztucznej inteligencji potwierdziły te ustalenia. Specjaliści często przedkładali interesy firmy nad sprawiedliwość, podkreślając potrzebę jaśniejszych wytycznych dotyczących tego, kiedy należy pominąć sugestie AI.

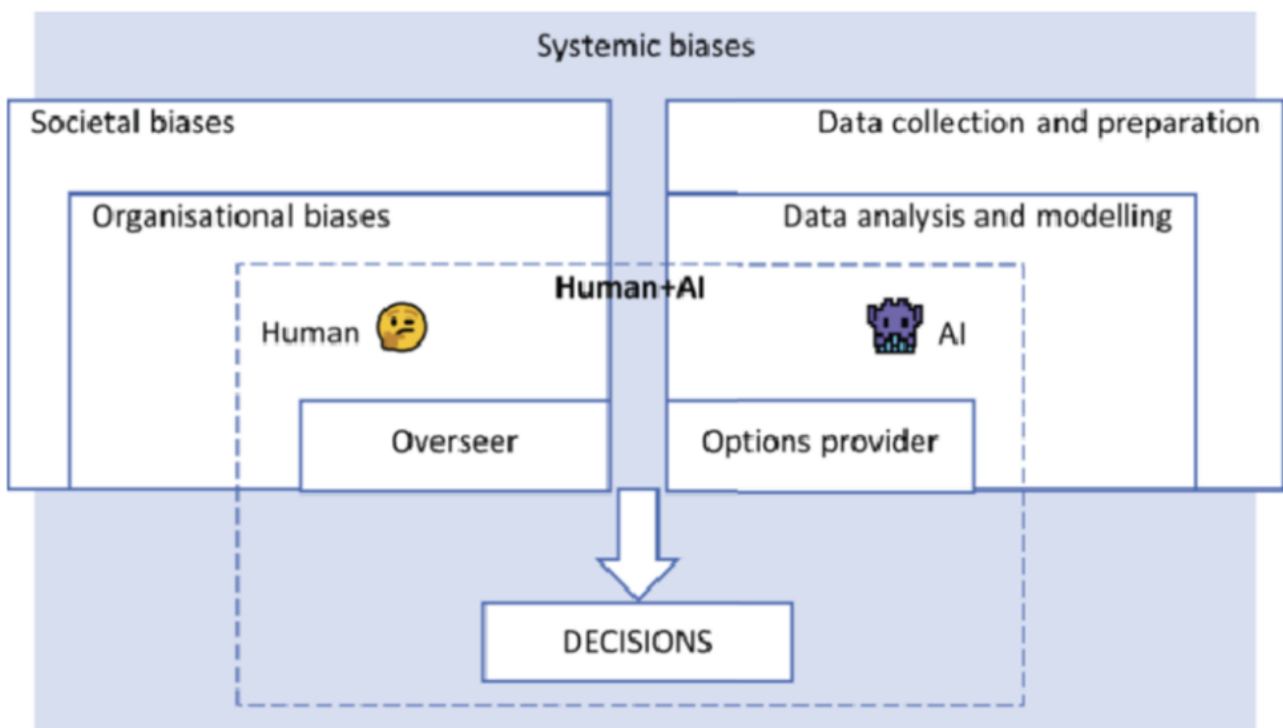
## W kierunku sprawiedliwości systemowej

Wnioski z badania wskazują na potrzebę przejścia od indywidualnego nadzoru do zintegrowanego systemu zaprojektowanego w celu przeciwdziałania uprzedzeniom zarówno ze strony człowieka, jak i sztucznej inteligencji. Prawdziwy nadzór obejmuje coś więcej niż

## Zrozumienie wpływu interakcji człowieka ze sztuczną inteligencją na dyskryminację

tylko programowanie sztucznej inteligencji tak, aby była sprawiedliwa lub poleganie na indywidualnym osądzie. Wymaga on holistycznego podejścia, takiego jak

- Środki techniczne mające na celu zapewnienie, że systemy sztucznej inteligencji są projektowane i aktualizowane z myślą o sprawiedliwości.
- Strategie organizacyjne mające na celu promowanie kultury, która priorytetowo traktuje sprawiedliwość i zapewnia szkolenia w zakresie zarządzania narzędziami sztucznej inteligencji.
- Interwencje polityczne, które ustanawiają jasne wytyczne dotyczące współpracy człowieka ze sztuczną inteligencją.



*Tło decyzyjne człowieka i sztucznej inteligencji oraz ich wzajemne oddziaływanie (Komisja Europejska)*

# Wzmocnienie pozycji decydentów

Aby skutecznie łagodzić uprzedzenia, decydenci potrzebują narzędzi i wytycznych, które pomogą im zrozumieć, kiedy i jak zastąpić zalecenia AI. Ciągłe monitorowanie i ocena wyników wspomaganych przez sztuczną inteligencję są niezbędne do identyfikowania i eliminowania pojawiających się uprzedzeń. Co więcej, dając decydentom dostęp do danych na temat ich wyników i potencjalnych uprzedzeń, możemy wspierać bardziej refleksyjne i odpowiedzialne podejście do podejmowania decyzji wspieranych przez sztuczną inteligencję.

# Informowanie o akcie o sztucznej inteligencji i nie tylko

Na początku ubiegłego roku przyjęto unijny akt prawny dotyczący sztucznej inteligencji, wyznaczający standardy regulacji w tym zakresie na całym świecie. Ustalenia KE są szczególnie istotne dla wprowadzenia tego aktu w życie, podkreślając konsekwencje ludzkiego nadzoru. Dostarczając praktycznych spostrzeżeń, Komisja stara się informować o przyszłych standardach i wytycznych, które nie tylko zapewnią zgodność z ustawą o sztucznej inteligencji, ale także ujawnią praktyczne kwestie wdrożeniowe, które są równie ważne. Zapraszamy do zapoznania się z raportem Komisji Europejskiej, oferującym bardziej szczegółowy opis badania i jego implikacji dla przyszłości zarządzania sztuczną inteligencją.

## Zrozumienie wpływu interakcji człowieka ze sztuczną inteligencją na dyskryminację

Przeczytaj raport

Dołącz do Komisji w przemyśleniu roli ludzkiego nadzoru w sztucznej inteligencji. Dzięki odpowiednim regulacjom i wytycznym możemy upewnić się, że UE może czerpać korzyści z nowych technologii, zapewniając jednocześnie sprawiedliwe i odpowiedzialne podejmowanie decyzji w erze cyfrowej. Jeśli chcesz dowiedzieć się więcej o tym projekcie, zapoznaj się z niektórymi zastosowanymi w nim metodami: Uczciwe podejmowanie decyzji: Czy ludzie mogą uchronić nas przed stronniczą sztuczną inteligencją? - Komisja Europejska

# Źródło:

artykuł Komisji Europejskiej