

Halucynacje, dyskryminacja, wyciek danych – rozmawiamy o AI w urzędzie



Stosunkowo niedawno znajomy informatyk powiedział mi, że jak zrobił analizę ruchu sieciowego to okazało się, że większość pracowników sporej firmy, w której pracował, codziennie korzystała z narzędzi określanych jako sztuczna inteligencja. Oczywiście korzystając z kont prywatnych, bo firma nie kupiła takich narzędzi dla siebie. Myśli Pan, że w samorządach byłoby podobnie?

Niestety, ale moje doświadczenia pokazują, że dokładnie tak jest. Wielu pracowników próbuje eksperymentować z nową technologią. Nie uważam tego za błąd, ponieważ rozwój napędzany jest poprzez potrzebę ułatwiania sobie życia i pracy. Każdy z kierowników JST powinien jednak zdać sobie sprawę z tego, że brak nadzoru nad wykorzystaniem takich narzędzi może prowadzić do bardzo dużych problemów. Problemy z ochroną danych osobowych, halucynacje czy dyskryminacja przez algorytmy to jedna z najpoważniejszych.

Dlaczego akurat ochrona danych osobowych jest tak wrażliwa jeżeli chodzi o korzystanie ze sztucznej inteligencji?

Każdy z nas korzystając z usług publicznych jest przekonany, że jego dane są bezpieczne i przetwarzane zgodnie z prawem. Nieumiejętne wykorzystanie narzędzi w szczególności niezgodnie z postanowieniami regulaminów może prowadzić do nieuprawnionej utraty ich poufności, integralności lub dostępności. Pracownik nieświadomie może udostępnić dane osobowe mieszkańców na podstawie, których algorytm będzie trenowany, jest to oczywiście niezgodne z przepisami, gdyż taki cel nie został mieszkańcom przedstawiony, ponadto dane takie mogą zostać upublicznione w ramach innych zapytań przez innych użytkowników. Musimy pamiętać, że przechowywanie danych tylko w pamięci narzędzi dostawców usług AI może wiązać się również z ich utratą, w szczególności jeżeli zasoby te są po za naszą infrastrukturą.

Jakie są inne zagrożenia związane z AI w naszej codziennej pracy?

Halucynacje i dyskryminacja, to dwa hasła powinny zawsze przyświecać użytkownikom narzędzi LLMowych ponieważ występują bardzo często. Halucynacje to uzyskanie odpowiedzi brzmiącej bardzo wiarygodnie, ale całkowicie lub częściowo mijającą się z prawdą. Natomiast dyskryminacja to zbudowanie przez algorytm własnych powiązań z danymi mu przekazanych, krótko opowiadając algorytm sam może zdecydować np. że w większości decyzji administracyjnych wydawanych np. osobom niepełnosprawnym na wycinkę drzewa była negatywna i podpowiadać, iż każda osoba niepełnosprawna powinna otrzymać decyzję negatywną. Mimo, że ta cecha osoby fizycznej nie powinna być wcale brana pod uwagę.

W jakich obszarach w samorządach widzi Pan najwięcej pól do pracy z AI i dlaczego? Jakież przykłady?

Praca z dokumentami to w mojej ocenie najbliższe zastosowanie modeli językowych, jest tu wiele obszarów do zastosowania od automatycznej dekretacji po propozycje pism czy wsparcie w zakresie archiwizacji. Oczywiście nie ma tutaj ograniczeń poza wyobraźnią, jednak każdy samorząd w pierwszej kolejności powinien skupić się na efekcie jaki chce osiągnąć, a dopiero na tym w jaki sposób go zrealizować, gdyż AI nie powinniśmy traktować jako jakąś nadprzyrodzoną moc, a narzędzie do wykorzystania

Czy widzi Pan wydziały w samorządach, które spokojnie mogą takich narzędzi używać bez specjalnych ograniczeń? Ja sobie wyobrażam np. wydział promocji. Z resztą nie wyobrażam sobie, by jakiś wydział promocji z tego nie korzystał, nawet jeśli formalnie JST nie ma polityk, czy zasad korzystania.

Również uważam wydział promocji za dobry przykład, w szczególności jeżeli ich działalność z modelami zamyka się tylko do generowania obrazów, filmów lub korekty tekstu. Powinniśmy tylko pamiętać o transparentnym komunikowaniu, iż dany utwór powstał przy wykorzystaniu AI.

Korzystając z okazji chciałbym zaprosić wszystkie samorządy i ich jednostki organizacyjne do wsparcia działań Prezesa Urzędu Ochrony Danych Osobowych Pana Mirosława Wróblewskiego w zakresie przygotowania podmiotów publicznych do bezpiecznego korzystania z AI poprzez wypełnienie krótkiej ankiety dostępnej na stronie UODO:

<https://uodo.gov.pl/pl/138/3929>

Ankieta została przygotowana we współpracy z wybitnymi specjalistami w dziedzinie prawa nowych technologii czyli Społecznym Zespołem Ekspertów przy Prezesie UODO i działającą w jego ramach Grupą roboczą ds. Sztucznej Inteligencji, pozwoli ona ukształtować przyszłość wykorzystania AI w administracji publicznej.

Ankietę można wypełnić do 15 listopada 2025.

Z Mariuszem Stasiakiem vel Staskiem, Inspektorem Ochrony Danych Osobowych w NIST rozmawiał Andrzej Gniadkowski.

Dziękujemy za tak liczne uczestniczenie w szkoleniach NIST i zapraszamy do śledzenia naszej oferty w zakładce "Szkolenia" (można też kliknąć w obrazek poniżej):