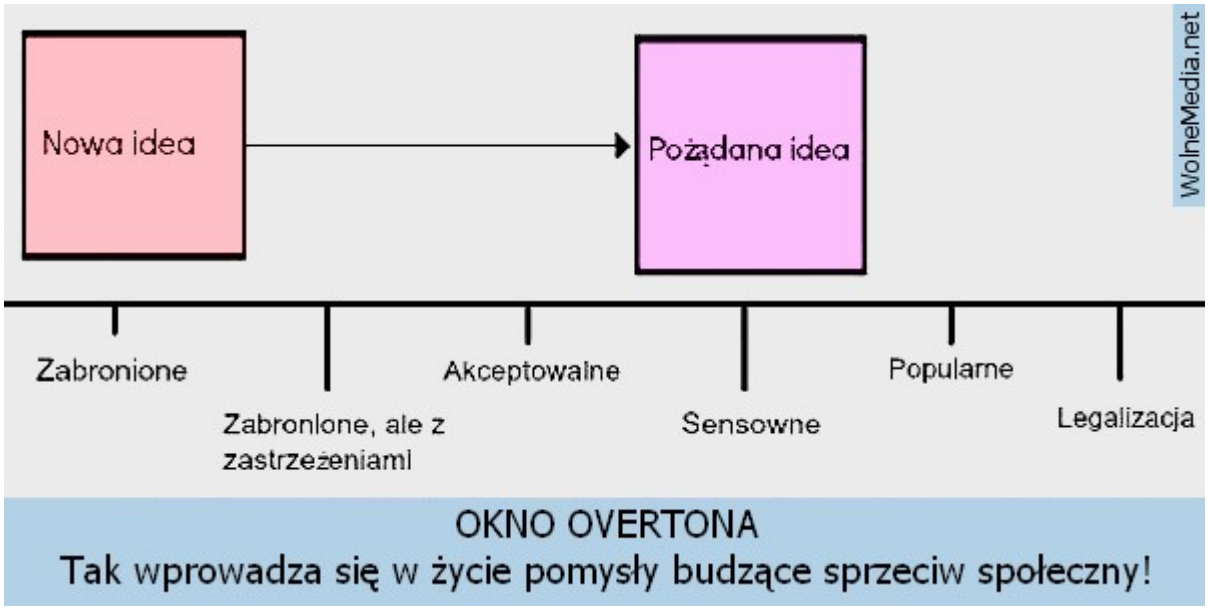
Profilowanie i ocenianie ludzi dzięki sztucznej inteligencji

22 czerwca 2024

Profilowanie i ocenianie ludzi pod kątem ich wiarygodności, przydatności, wypłacalności to jedno z bardziej ryzykownych zastosowań sztucznej inteligencji. Co mówi o nim „AI Act”?



W tego rodzaju zastosowaniach zadaniem AI jest przewidzieć zachowanie albo predyspozycje konkretnej osoby na podstawie jej danych, a także na podstawie wiedzy o tym, jak zachowywały się w przeszłości osoby o podobnych cechach. Chociaż zanim akt o sztucznej inteligencji zacznie w pełni obowiązywać, minie wiele miesięcy (a nawet lat, bo niektóre jego przepisy wejdą w życie dopiero w roku 2030), już teraz przyglądamy się, co nowe prawo może zmienić w relacji między ludźmi a narzędziami wykorzystującymi uczenie maszynowe. W tym odcinku analizujemy systemy profilujące ludzi.

Co to są systemy profilujące i do czego służą

Jeśli czytaliście „Raport mniejszości” Phillipa K. Dicka lub oglądaliście nakręcony na podstawie tej powieści film Stevena Spielberga, to pewnie domyślacie się, czym jest predykcja kryminalna. W dużym uproszczeniu: jest to przewidywanie, kto popełni przestępstwo. W rzeczywistości takie systemy działają m.in. w Stanach Zjednoczonych (np. COMPAS wykorzystywany do szacowania ryzyka ponownego popełnienia przestępstwa przez osobę osadzoną) czy w Holandii (Top 400 – lista młodych ludzi „wysokiego ryzyka”).

„W Amsterdamie w Holandii działa narzędzie opracowane przez władze miejskie w porozumieniu z policją i prokuraturą. Opiera się na profilowaniu młodzieży poniżej 16 roku życia. W ramach tego systemu powstaje lista obejmująca „top 400” osób wysokiego ryzyka. I gdy w okolicy dochodzi do przestępstwa, to właśnie nazwiska z tej listy mogą znaleźć się jako pierwsze w kolejce do nadzoru, przesłuchań, przeszukań i aresztowań. A rodzicom dzieci z takiej listy można zagrozić odebraniem praw rodzicielskich. To tylko jeden przykład. (…) Do listy krajów stosujących podobne narzędzia można dopisać także Hiszpanię. Tam istnieje między innymi system do oceny ryzyka ekstremizmu. I on szybko przekształcił się w narzędzie do śledzenia muzułmanów, bo tak go zaprojektowano” – powiedział Griff Ferris w rozmowie z Jakubem Dymkiem.

Chociaż automatyzacja w założeniu ma prowadzić do wyeliminowania uprzedzeń, praktyka pokazuje, że dzieje się odwrotnie. Kolejny przykład z Holandii: według Lighthouse Reports algorytm profilowania wykorzystywany do oceny ryzyka w stosunku do osób aplikujących o wizę lub pobyt krótkoterminowy w Holandii i strefie Schengen bazuje na zmiennych takich jak narodowość, wiek i płeć. Statystyki pokazują, że do kategorii „wysokiego ryzyka” nieproporcjonalnie często trafiają Surinamczycy w wieku 26–40 lat, którzy aplikowali z

Paramaribo, a także nieżonaci Nepalczycy w wieku 35–40 lat, którzy aplikowali o wizy turystyczne. 33% wniosków osób z kategorii „wysokiego ryzyka” jest odrzucanych (przy 3,5% odmów dla osób z kategorii „niskiego ryzyka”). System pozostaje w użyciu – i to wbrew sprzeciwom ze strony inspektora ochrony danych w holenderskim Ministerstwie Spraw Zagranicznych, który podkreśla, że jego stosowanie wiąże się z dyskryminacją ze względu na pochodzenie.

Podobny system działał w Wielkiej Brytanii, gdzie Ministerstwo Spraw Wewnętrznych przy rozpatrywaniu wniosków wizowych posiłkowało się tajną listą „podejrzanych” narodowości. Osoba, która należała do jednej z nich, z automatu dostawała czerwoną flagę, co najczęściej było równoznaczne z odmową przyznania wizy. Po wniesieniu sprawy do sądu przez organizacje działające na rzecz praw człowieka ministerstwo zapowiedziało, że „dokładnie sprawdzi działanie systemu, także pod kątem potencjalnego »niezamierzonego« odchylenia lub dyskryminacji”.

Systemy profilujące ludzi wykorzystywane są też w celach biznesowych. Firmy używają ich np. do przesiewania setek dokumentów składanych w ramach rekrutacji ubezpieczyciele – do szacowania ryzyka, banki – w modelach scoringowych. Coraz odważniej sięgają też po nie instytucje publiczne – do typowania wyłudzeń. Profilowanie jest też wykorzystywane w marketingu do personalizowania reklam i konkretnych ofert.

Cele mogą być bardzo różne, ale zasada działania jest podobna: systemy profilujące ludzi porównują dane zebrane na temat konkretnej osoby z ustrukturyzowaną wiedzą o „podobnych przypadkach” (na którą składają się wcześniejsze obserwacje poczynione na dużej próbie statystycznej), żeby wydedukować brakujące elementy układanki (np. jaka jest wiarygodność kredytowa albo przygotowanie do pracy danej osoby) lub przewidzieć jej zachowanie w przyszłości. Takie wnioskowanie pozwala podjąć decyzję w konkretnej sprawie mimo braku pełnych danych, czasem nawet bez udziału człowieka. W teorii ma to służyć zwiększeniu efektywności podejmowanych decyzji i wymóc

obiektywizm. Jak jest w praktyce – o tym za chwilę.

Różne poziomy ryzyka – różne obowiązki, czyli co zmieni „AI Act”

W „AI Act” nie znajdziemy definicji systemów profilujących. W tym zakresie regulacja odsyła do RODO:

– Art. 3 pkt 52 AIA: „Profilowanie” oznacza profilowanie zdefiniowane w art. 4 pkt 4 rozporządzenia (UE) 2016/679 [RODO];

– Art. 4 ust. 4 RODO: „Profilowanie” oznacza dowolną formę zautomatyzowanego przetwarzania danych osobowych, które polega na wykorzystaniu danych osobowych do oceny niektórych czynników osobowych osoby fizycznej, w szczególności do analizy lub prognozy aspektów dotyczących efektów pracy tej osoby fizycznej, jej sytuacji ekonomicznej, zdrowia, osobistych preferencji, zainteresowań, wiarygodności, zachowania, lokalizacji lub przemieszczania się.

„AI Act” natomiast dzieli systemy AI na różne kategorie – w zależności od wpływu, jaki mogą one wywierać na prawa człowieka. Pierwsza kategoria to systemy zakazane. Do niej zaliczone zostaną systemy profilujące, których użycie prowadzi do krzywdzącego lub niekorzystnego traktowania (takie jak chiński SCS), i systemy, które służą do oceny ryzyka popełnienia przestępstwa (jak amerykański COMPAS i holenderski Top 400).

Druga, znacznie pojemniejsza, kategoria to systemy wysokiego ryzyka. Zgodnie z „AI Act” do tej kategorii powinien zostać zaliczony każdy system AI, którego działanie wiąże się z profilowaniem osób fizycznych (niezależnie od kontekstu jego zastosowania). Komisja Europejska ma opracować listę przykładowych systemów profilujących. Typowe zastosowania takich systemów opisaliśmy powyżej.

Dzięki zawartym w „AI Act” procedurom (szczegółowo opisanym w ramce) błędy i odchylenia algorytmów powinny być wyłapywane, a potem naprawiane, zanim system trafi na rynek. Takie błędy również nie powinny się pojawiać na późniejszym etapie, bo podmiot wdrażający (np. firma prowadząca rekrutację przy wsparciu systemu AI) musi zadbać o to, żeby dane wejściowe były „istotne i wystarczająco reprezentatywne w świetle zamierzonego celu”.

Dla systemów wysokiego ryzyka – wysokie standardy

Podmioty rozwijające systemy należące do kategorii wysokiego ryzyka będą musiały:

– zarządzić związanym z działaniem tych systemów ryzykiem np. dla naszego zdrowia, bezpieczeństwa i praw podstawowych (zidentyfikować ryzyko oraz skutecznie mu zapobiegać);

– testować systemy „w celu określenia najodpowiedniejszych i najbardziej ukierunkowanych środków zarządzania ryzykiem”;

– wprowadzić wewnętrzną kontrolę jakości, której elementem jest autentyczny i stały nadzór człowieka;

– stosować praktyki zarządzania danymi „odpowiednie do zamierzonego celu systemu sztucznej inteligencji wysokiego ryzyka” (przykłady poniżej);

– stworzyć „ramy odpowiedzialności” określające obowiązki kierownictwa i innego personelu – po to, żeby procedury nie zostały na papierze, tylko były autentycznie wdrażanie i przestrzegane w firmie. A jeśli nie są, żeby odpowiadał za to ktoś z zarządu.

Zgodnie z art. 14 „AI Act” systemy sztucznej inteligencji wysokiego ryzyka mają być projektowane i rozwijane w taki sposób, by „osoby fizyczne mogły je skutecznie nadzorować w

całym okresie ich użytkowania” (w tym przy użyciu odpowiednich interfejsów człowiek–maszyna).

W art. 10 „AI Act” podane zostały przykłady „odpowiednich do zamierzonego celu” praktyk zarządzania danymi:

– etykietowanie, czyszczenie, aktualizowanie, wzbogacanie i agregowanie danych;

– sformułowanie założeń „zapisanych” w danych;

– ocena dostępności, ilości i przydatności danych;

– badanie pod kątem możliwych stronniczości, które mogą mieć wpływ na zdrowie i bezpieczeństwo osób, mieć negatywny wpływ na prawa podstawowe lub prowadzić do dyskryminacji zakazanej na mocy prawa Unii, szczególnie w przypadku gdy uzyskane dane mają wpływ na wkład w przyszłe operacje;

– wykrywanie ewentualnych błędów i uprzedzeń, zapobieganie im lub łagodzenie ich;

– wykrywanie luk i braków w danych.

Ale są też wyjątki…

„AI Act” robi wyjątek od tego standardu dla systemów AI, które co prawda są wykorzystywane w obszarach wysokiego ryzyka, ale tylko do prostych zadań proceduralnych (np. przekształcanie danych nieustrukturyzowanych w ustrukturyzowane) lub wspomagania człowieka (bez zastępowania ludzkiej oceny). Z kategorii systemów wysokiego ryzyka są też wyłączone takie systemy, których zadaniem jest „wykrywanie wzorców podejmowania decyzji lub odstępstw od wzorców podjętych uprzednio decyzji”, pod warunkiem że rozstrzygnięcie systemu nie zastępuje oceny dokonywanej przez człowieka ani nie wywiera na nią wpływu.

Na papierze poprzeczka dla firm rozwijających i wdrażających

systemy wysokiego ryzyka jest więc zawieszona naprawdę wysoko. Jak je potraktują? Czy przestraszą się (naprawdę wysokich!) kar? A może zmotywuje je wizja „doskonałej sztucznej inteligencji", którą Unia Europejska wspiera, regulując rynek, i chce promować na świecie? Niedługo się przekonamy.

Sztuczna inteligencja analizuje CV. Co robić w razie dyskryminacji

A co powiecie na rozmowę rekrutacyjną online z chatbotem? Kate Bussmann w reportażu How to nail a job interview with a robot (yes, these exist) dowodzi, że może ona się skończyć zaskakującymi rozstrzygnięciami. Bussmann opisuje m.in. niemiecki eksperyment dziennikarski z 2021 r. Pokazał on, że w rozmowie o pracę prowadzonej na wideo lepiej oceniane były osoby, które siedziały na tle półek z książkami i obrazów, a także te z zakrytą głową. Gorzej – osoby w okularach.

„Oprogramowanie z elementami AI jest wykorzystywane w procesie rekrutacji, przeważnie do wyłapywania słów kluczowych w CV, żeby z dużej puli zgłoszeń wyłonić te „lepiej rokujące". Ale pojawiają się też systemy analizujące zachowanie i wygląd kandydatów w ramach wstępnej selekcji. Po taki sięgnęła m.in. firma Unilever: Specjalny algorytm skanuje i analizuje twarz kandydata do pracy, jego sposób wysławiania się, gestykulację, ton głosu podczas odpowiedzi na zestaw specjalnie przygotowanych pytań. Analiza dokonywana jest podczas wstępnej selekcji kandydatów, która odbywa się online" – czytamy w artykule „AI wskazuje kandydatów do pracy w korporacji. Są kontrowersje" opublikowanym dzienniku „Rzeczpospolita".

W sieci znajdziecie poradniki, jak przygotować się do rozmowy z robotem i jak przygotować CV, które spodobają się algorytmom. Ale być może nie będą już one potrzebne: dzięki starym przepisom RODO i nowym uprawnieniom wynikającym z „AI Act" mamy narzędzia, by bronić się przed dyskryminującym albo błędnym wyrokiem sztucznej inteligencji.

Jak RODO i „AI Act” chronią nas przed dyskryminującym albo błędnym wyrokiem sztucznej inteligencji

Konieczność podania informacji o tym, że podlegamy ocenie systemu sztucznej inteligencji, którą można potraktować jako ostrzeżenie: „Podmioty wdrażające systemy sztucznej inteligencji wysokiego ryzyka, które podejmują decyzje lub pomagają w podejmowaniu decyzji dotyczących osób fizycznych, informują osoby fizyczne, że podlegają korzystaniu z systemów sztucznej inteligencji wysokiego ryzyka” [art 26 ust. 11 „AI Act”].

Prawo do niepodlegania „w pełni zautomatyzowanej” decyzji i prawo do ludzkiej interwencji: „Artykuł 22 zakazuje podejmowania decyzji wyłącznie na podstawie automatycznego przetwarzania danych, włączając w to profilowanie, „jeżeli te decyzje mają prawne skutki względem osoby lub w podobny sposób istotnie na nią wpływają”. Oznacza to, że nie można używać tylko sztucznej inteligencji do podejmowania ważnych decyzji, takich jak przyznanie kredytu, zatrudnienie czy znaczące decyzje medyczne – bez ludzkiego nadzoru”. Od tego zakazu są wyjątki, ale nawet gdy automatyczne podejmowanie decyzji jest dozwolone, istnieje prawo do interwencji ludzkiej. Oznacza to, że można domagać się, aby ostateczną decyzję rozpatrzył człowiek, a nie tylko AI.

Prawo do wyjaśnienia podstaw rozstrzygnięcia podjętego przy użyciu systemu AI wysokiego ryzyka oraz roli systemu AI w procesie podejmowania decyzji [art. 86 „AI Act”]: „Każda osoba, na którą AI ma wpływ, będąca przedmiotem decyzji podjętej przez podmiot stosujący na podstawie wyników systemu AI wysokiego ryzyka wymienionego w załączniku III, z wyjątkiem systemów wymienionych w pkt 2 tego załącznika, która to decyzja wywołuje skutki prawne lub w podobny sposób oddziałuje na tę osobę na tyle znacząco, że uważa ona, iż ma to

niepożądany wpływ na jej zdrowie, bezpieczeństwo lub prawa podstawowe, ma prawo uzyskania od podmiotu stosującego jasnego i merytorycznego wyjaśnienia roli tego systemu AI w procedurze podejmowania decyzji oraz głównych elementów podjętej decyzji" [art. 86 „AI Act"].

Prawo do złożenia skargi na źle działający (np. krzywdzący) system AI do organu monitorującego rynek (tj. sprawdzającego, czy systemy AI w Polsce działają zgodnie z prawem) [art. 68a „AI Act"]. Skargę można złożyć na podstawie wyjaśnienia, o którym piszemy w punkcie wyżej.

Prawo dochodzenia odszkodowania lub zadośćuczynienia – w obszarze rekrutacji na podstawie art. 18 (3a) §1 kodeksu pracy, w innych przypadkach – na podstawie kodeksu cywilnego.

Trudno powiedzieć, jak te uprawnienia zadziałają w praktyce. Ale dobrze wiedzieć, że te ścieżki istnieją.

AI w służbie państwa. Czy wyciągnęliśmy wnioski ze skandalu z SyRI?

Pewnego dnia Chermaine odebrała list z urzędu skarbowego, w którym żądano od niej zwrotu zasiłków, które państwo wypłacało jej na opiekę nad dziećmi przez poprzednie cztery lata. Kwota była niebagatelna: 100 tys. euro. Chermaine, wówczas matka trójki małych dzieci i jeszcze studentka, pomyślała, że to błąd. Ale to nie był błąd. To był początek maratonu przypominającego Proces Kafki.

„Skandal zasiłkowy", którego ofiarą padła ta młoda kobieta, rozegrał się w Holandii w 2012 r. Dopiero 7 lat później wyszło na jaw, że holenderski fiskus określał ryzyko nadużyć w systemie przy pomocy algorytmu uczenia maszynowego. Dziesiątki tysięcy rodzin, w większości niezamożnych i należących do mniejszości etnicznych, z dnia na dzień zostało bez środków do

życia oraz z ogromnymi długami. I to nie dlatego, że rzeczywiście wyłudziły zasiłek… System wytypował ich jako podejrzanych, którzy mogliby dopuścić się takiego czynu.

W przypadku Chermaine skończyło się „tylko" na rozstaniu z partnerem, depresji i wypaleniu, ale wiele osób dotkniętych skandalem popełniło samobójstwo. Ponad tysiąc dzieci zostało odebranych rodzicom i trafiło do rodzin zastępczych. Sprawa odbiła się głośnym echem w międzynarodowej prasie i wywołała kryzys polityczny w Holandii. W 2020 r. sąd okręgowy w Hadze uznał, że przepisy, na podstawie których działał system SyRI, naruszają Europejską konwencję praw człowieka.

Obietnica większej efektywności i nieomylności sztucznej inteligencji kusi rządy i instytucje kolejnych państw. W tym polski ZUS, który już używa algorytmów do przetwarzania wniosków o świadczenie wychowawcze (tzw. 800+). Czy wysokie standardy, jakie dla systemów wysokiego ryzyka przewiduje „AI Act", zabezpieczą nas przed błędami twórców holenderskiej SyRI? Czy zadziała kontrola jakości i nadzór człowieka? Oby.

Autorstwo: Anna Obem Anna, Katarzyna Szymielewicz
Konsultacja: dr Gabriela Bar (Gabriela Bar Law & AI)
Źródło: Panoptykon.org

Komentarz „Wolnych Mediów"

Profilowanie ludzi przez AI pod kątem możliwych zachowań i reakcji społecznych jest tzw. elitom niezbędne do wprowadzenia cyfrowego totalitaryzmu NWO po Wielkim Resecie. Krok po kroku ludzie będą profilowani, nękani wezwaniami do wyjaśnień, zastarszani podczas nich, a nawet zapobiegawczo karani za niepopełnione czyny (np. odbieraniem różnych praw, np. do przemieszczania się lub korzystania z jakichś usług, które mogą zwiększyć pradopodbienstwo popełnienia czynu szkodliwego społecznie). Teraz do tego dodajmy chińską ideę oceny społecznej, limity śladu węglowego, pieniądze ze wskazaniem celu wydatkowego, itd. i mamy totalitaryzm i masowe

niewolnictwo, o którym nie śniło się nawet Orwellowi. Nietrudno sobie wyobrazić, że władze wykorzystają w przyszłości profilowanie przez AI do utrzymania swej władzy i prześladowania potencjalnej opozycji. Np. osobom podejrzanym o sprzyjanie spotanicznym protestom zakażą przebywania w miejscach protestogennych, albo odetną dostęp do stron internetowych krytykujących rząd, by ich poglądy nie wyewoluowały w jakiś bunt lub rewolucję.