Biuletyn Fundacji Panoptykon nr 4

(O)BŁĘDNA INTELIGENCJA

W BIULETYNIE:

REDAKCJA: Małgorzata Szumańska

WYWIADY: Jakub Dymek

KOORDYNACJA WYDANIA: Justyna Dywańska

PROJEKT GRAFICZNY I SKŁAD: Olek Modzelewski

KOREKTA: Urszula Dobrzańska

ISBN: 978-83-966914-1-5

Fundacja Panoptykon
Warszawa 2024

Papier użyty do druku biuletynu w 100% pochodzi z makulatury i został wyprodukowany z troską o minimalizowanie negatywnego wpływu na środowisko.

FUNDACJA PANOPTYKON

Przekłujmy ten balonik

Trudno to przeoczyć: świat oszalał na punkcie sztucznej inteligencji. I w tym roku właśnie jej – a jakże! – poświęcamy biuletyn Panoptykonu. Ale nie po to, by dmuchać w ten napięty do granic wytrzymałości balonik, tylko po to, by spróbować go przebić.

Ekscytacja AI (ang. *artificial intelligence*) utrudnia zauważenie, że to po prostu narzędzie. Fakt – w niczym nie przypomina noża. Zarówno możliwości, jakie daje, jak i zagrożenia, które wywołuje, są nieporównywalne do ryzyka związanego z korzystaniem z najostrzejszego nawet narzędzia do krojenia chleba. Czy pozwoli zbudować bardziej racjonalne i sprawiedliwe społeczeństwo? Czy może przejmie kontrolę nad światem, a nas wszystkich unicestwi? Taka dyskusja prowadzi donikąd.

Sztuczna inteligencja działa tu i teraz: zmienia świat wokół i nas samych, nie tylko generując zyski czy – nieraz pozorne – oszczędności, ale również tworząc szereg konkretnych problemów. Z każdym rokiem przybywa ryzykownych pomysłów na jej wykorzystanie. Szum robi się wtedy, kiedy wychodzi na jaw, że może się mylić i powielać stereotypy.

Weźmy system, który ma typować osoby wyłudzające zasiłki. Nakarmiony pełnymi uprzedzeń danymi, popełnia błędy, które mają dramatyczne konsekwencje: dzieci trafiają do domów dziecka, dorośli popełniają samobójstwa (SyRI). Albo system, który na podstawie analizy mikrogestów osoby ubiegającej się o prawo wjazdu do Unii Europejskiej ocenia, czy ta mówi prawdę (iBorderCtrl). Czy będzie ślepy na kolor skóry i status społeczny? Albo algorytm, który „skanuje i analizuje twarz kandydata do pracy, jego sposób wysławiania się, gestykulację, ton głosu podczas odpowiedzi na zestaw specjalnie przygotowanych pytań" (firma Unilever). Czy da równe szanse kobietom i mężczyznom?

Systemy oparte na AI wyłapują prawidłowości w zachowaniu ludzi i na tej podstawie wyciągają wnioski. Nie zawsze słuszne. Błędy skrywają historie osób, którym przyszło mierzyć się z jakąś krzywdzącą decyzją. Jednak to nie sztuczna inteligencja jest odpowiedzialna za ich dramaty, lecz ludzie, którzy decydują, jakie zadania na nią delegować, na jakich danych ją trenować, jakie cele przed nią postawić. To ich poczynaniom powinniśmy się przyglądać, a nie temu, co „robi" AI (bo samo jeszcze niczego nie robi). Niestety, oddalają nas od tego popularne wyobrażenia sztucznej inteligencji jako wszechpotężnego supermózgu. Jeśli – jak przekonuje Cathy O'Neil – uwierzymy w mit maszyny obdarzonej świadomością, nie będziemy zdolni jej kwestionować.

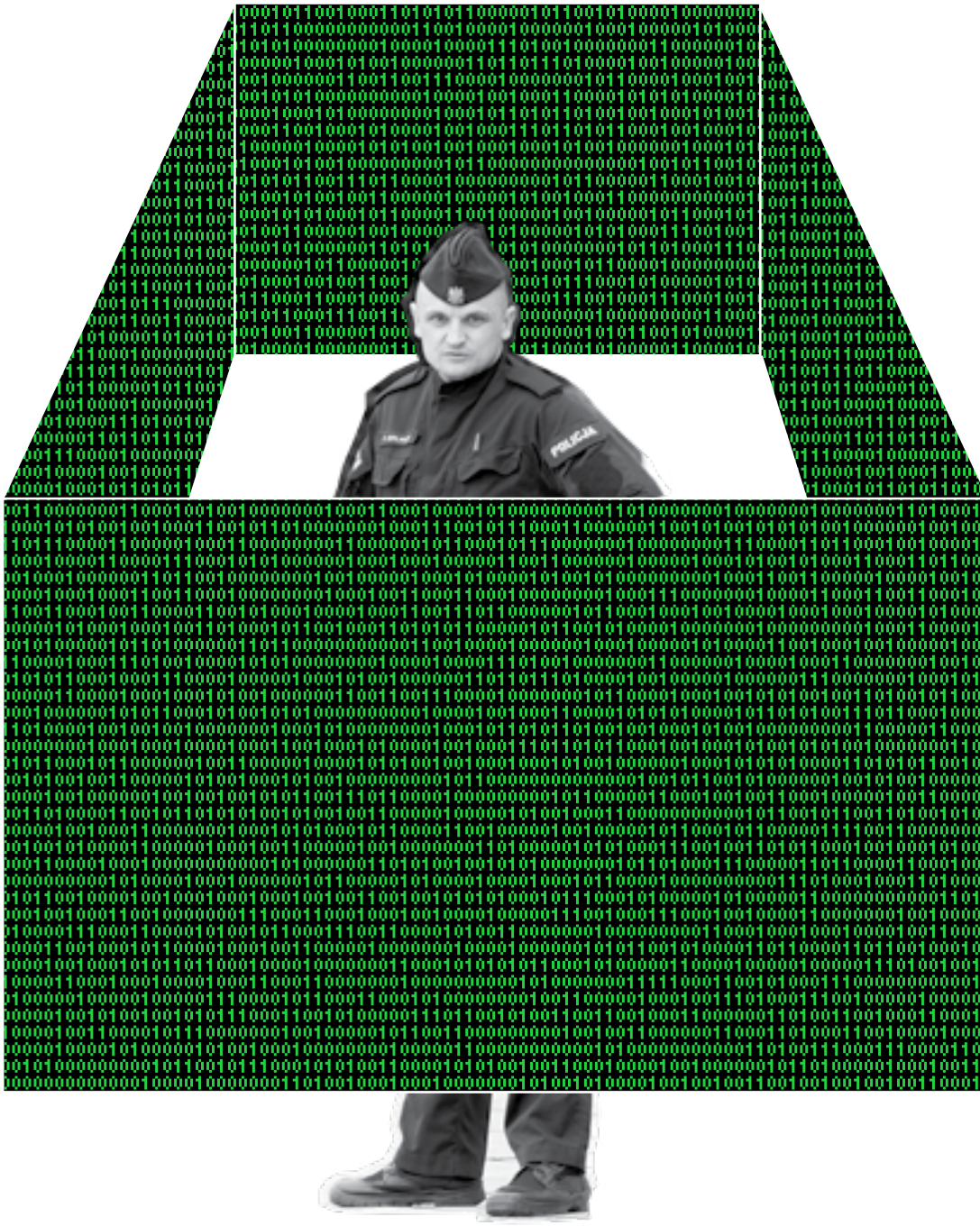
A gdyby tak się udało to „patrzenie na ręce" na kogoś delegować? Gdyby jakiś organ kontrolował, czy tworzący i wdrażający AI przestrzegają ustalonych standardów? A systemy przed wejściem na rynek przechodziłyby „testy bezpieczeństwa"?

Taki był pomysł Komisji Europejskiej na regulacje: nie ograniczać innowacji, ale podnieść poprzeczkę tam, gdzie błędy mogą mieć szczególnie dotkliwe skutki. Niestety, przyjęte w tym roku rozporządzenie o sztucznej inteligencji (AI Act) zatrzymuje się w pół drogi. Sukcesem ruchu – w którym działa Panoptykon – jest wprowadzenie ludzkiej perspektywy do europejskiej debaty politycznej zdominowanej na początku przez biznes i jego cele. Ale nie oszukujmy się – ten akt prawny nie był pisany z myślą o ludziach, którzy są za pomocą AI profilowani, sortowani, nadzorowani i oceniani. W dodatku zawiera tyle luk, że jego egzekwowanie będzie prawdziwym wyzwaniem.

Czy w takim razie to prawo ma w ogóle szansę coś zmienić? Sporo zależy od tworzących i wdrażających AI (czy zaakceptują ograniczenia, czy będą eksploatować niejasności). Ale też np. od polskiego rządu (czy zrezygnuje z szerokich wyłączeń dla systemów służących bezpieczeństwu publicznemu, czy powoła silny organ kontrolny).

Jeśli to ty masz wpływ na rozwój, wykorzystanie i regulację AI, to chcemy ci coś przekazać. Rozwój sztucznej inteligencji nie powinien być celem samym w sobie. Nie potrzebujemy więcej technologii. Potrzebujemy technologii, która służy ludziom. Nie musi być nie wiadomo jak „inteligentna", ale za to tworzona z głową i mądrze wykorzystywana. Dlaczego? Właśnie o tym przeczytasz w biuletynie.

Katarzyna Szymielewicz
Małgorzata Szumańska


POLICJA

CYFROWY POLICJANT NAS NIE URATUJE

Rozmowa Jakuba Dymka z Griffem Ferrisem – badaczem wpływu technologii na prawa człowieka w wymiarze sprawiedliwości

To samo, co zrobiono z kasjerami w supermarketach, dzieje się powoli z policjantami i sędziami. Obiecano nam, że technologia rozwiąże wszystkie problemy, o ile my zgodzimy się wszystkie swoje problemy i troski powierzyć w ręce technologii.

Jakub Dymek: Gdy próbowałem zgłębić temat wpływu sztucznej inteligencji na wymiar sprawiedliwości, co krok napotykałem optymistyczne prognozy i scenariusze. „Sztuczna inteligencja przyniesie rewolucję w podejściu do bezpieczeństwa i uwolni wielkie zasoby po stronie policji" – to tylko jeden z wielu podobnych cytatów. Czytam, że policja dzięki zastosowaniu algorytmów będzie miała więcej czasu, biurokracja zostanie ograniczona, a automatyzacja i cyfryzacja przyspieszą działanie wymiaru sprawiedliwości. Gdzie zatem problem?

Griff Ferris: Cóż, technologiczny solucjonizm jest problemem sam w sobie. Widzimy to w kolejnych dziedzinach życia: w transporcie, ochronie zdrowia, świadczeniach społecznych, a obecnie w bezpieczeństwie publicznym i wymiarze sprawiedliwości. Za każdym razem słyszymy, że nowe technologie nas uratują, że sztuczna inteligencja zajmie się za nas różnymi patologiami. Teraz – że AI wyłapie przestępców, oczyści nasze ulice z chuligaństwa, wniesie cyfrową precyzję i neutralność do sądów… A nikt nie zadaje podstawowych pytań.

Na przykład jakich?

Co AI, czyli „sztuczna inteligencja", miałaby w tym kontekście oznaczać. Kiedy ktoś przekonuje nas o zbawiennych skutkach jakiejś technologii, a my nawet nie wiemy, jak ona działa – to czy nie powinno nas to zastanawiać? A dziś jest tak, że właściwie wszystko, co opiera się na jakiejś bazie danych, oprogramowaniu i nawet bardzo luźno pojętej automatyzacji, od razu nazywane jest sztuczną inteligencją. Gdy więc na przykład policja wprowadza rozwiązania bazujące na kartotekach czy informacjach, które już posiada, także nazywa to „wdrażaniem AI". Można mieć wrażenie, że rzeczywiście wchodzimy na inny poziom. To jednak złudne.

Dlaczego?

Z każdym narzędziem bazującym na przetwarzaniu danych przez algorytmy, a tym bardziej na uczeniu maszynowym, wiąże się zasadnicze ryzyko. Takie, że algorytm będzie powielał prawidłowości i uprzedzenia zawarte w tych danych i „uczył się" ich. Bazy danych odzwierciedlają bowiem strukturę władzy i mechanizmy działania tego społeczeństwa, które je stworzyło. W policyjnych kartotekach w państwach europejskich znajdziemy nieproporcjonalnie dużo ludzi biednych, o innym niż biały kolorze skóry czy Romów, osób marginalizowanych – historycznie zawsze tak było. To osoby o mniej pewnej pozycji w społeczeństwie były najczęściej obiektem nadzoru, przeszukań, aresztów i śledztw. Zatem organy ścigania zawsze dysponowały większą bazą danych na ich temat. W oczywisty sposób wprowadzenie tutaj rachunku algorytmów i „sztucznej inteligencji" będzie skutkowało brakiem neutralności systemów cyfrowych.

Rozumiem: wymiar sprawiedliwości historycznie nie był wolny od uprzedzeń. Ale co zmienia w tym AI?

Jeśli jesteśmy świadomi, że istnieje dyskryminacja w jakiejkolwiek formie, to chyba nie powinniśmy się zgadzać na jej dalsze trwanie, prawda? Dziś godzimy się nie tylko z istnieniem

tych uprzedzeń – rasizmu, seksizmu, homofobii – ale i na przeniesienie ich na poziom kolejnych algorytmów i technologicznych narzędzi w naszym wymiarze sprawiedliwości. Dopuszczamy myśl, że systemy, które będą wspomagać (a może nawet wyręczać) człowieka w fundamentalnych decyzjach, same będą dalekie od neutralności i obiektywizmu.

Ktoś może łatwo odpowiedzieć, że ryzyko błędu jest zawsze, ale wymiar sprawiedliwości w państwie demokratycznym gwarantuje prawo do obrony, odwołania się do wyższej instancji, domniemanie niewinności i wiele innych zabezpieczeń. Czy ryzyko, że padniemy ofiarą rasistowskiego czy seksistowskiego algorytmu, rzeczywiście jest aż tak duże?

Spójrzmy na to inaczej. Istnieje większe ryzyko, że nigdy nie dowiemy się, czy w naszej sprawie sięgnięto po jakieś narzędzia cyfrowe czy algorytmy – ani też w jaki sposób wpłynęły one na decyzję policji, prokuratury czy sądu. Dlaczego? Po pierwsze: władza nie ma ochoty ani obowiązku się z tego tłumaczyć. Po drugie: nawet jeśli wiemy, że nasze państwo korzysta z narzędzi sztucznej inteligencji, to firmy, które dostarczają te technologie, nie ujawniają, jak one działają. To ich pilnie strzeżony sekret: prawo do algorytmu i wiedza o jego funkcjonowaniu to często przecież najdroższe, co mają. Jeśli więc nie wiemy, co i jak jest wdrażane do działania organów ścigania i policji, to nie mamy się na co skarżyć i od czego odwoływać, prawda?

Ale to chyba najlepszy moment, żeby w końcu powiedzieć, jakich konkretnie systemów czy narzędzi powinniśmy się twoim zdaniem obawiać.

Jeśli chodzi o policję i wykorzystanie sztucznej inteligencji, mówimy dziś tak naprawdę o dwóch typach narzędzi: o algorytmach predykcyjnych i o ocenie ryzyka w odniesieniu do konkretnych osób. Te pierwsze przeczesują bazy danych policji i innych agencji, aby na tej podstawie (z uwzględnieniem aspektów geograficznych, urbanistycznych, demograficznych) wytypować miejsca, gdzie może dojść do łamania prawa. Te drugie zaś mają za zadanie oszacować, czy istnieje prawdopodobieństwo, że wytypowana osoba popełni w przyszłości przestępstwo.

I ktoś tego naprawdę używa?

Jest wiele różnych przykładów używania tego typu systemów. W Amsterdamie w Holandii działa narzędzie opracowane przez władze miejskie w porozumieniu z policją i prokuraturą, które bazuje na profilowaniu młodzieży poniżej 16 roku życia. W ramach tego systemu powstaje lista obejmująca „top 400" osób wysokiego ryzyka. Kiedy w okolicy dochodzi do przestępstwa, to właśnie nazwiska z tej listy mogą znaleźć się jako pierwsze w kolejce do przesłuchań, przeszukań i aresztowań. A rodzicom dzieci z takiej listy częściej grozi odebranie praw rodzicielskich. To tylko jeden z przykładów.

Takich rzeczy jest więcej?

Oczywiście.

W Wielkiej Brytanii używano systemu indywidualnej oceny ryzyka, którego wskazanie wpływało na prawdopodobieństwo postawienia zarzutów. To znaczy: jeśli ktoś mieścił się w profilu, który wskazywał na ryzyko złamania prawa w przyszłości, było bardziej prawdopodobne, że usłyszy zarzuty. Ale i na odwrót: osoby, które rzekomo popełniły jakieś wykroczenie, ale ich profil wskazywał na niskie ryzyko, mogły w ogóle nie usłyszeć zarzutów. Uznawano, że i bez tego się zresocjalizują i poprawią swoje postępowanie.

To jest przykład bardzo ważnych decyzji, które mogą wpłynąć na całe życie człowieka, a które w coraz większym stopniu powierzamy wyrokom AI.

Wybacz, że to powiem, ale bardzo to wszystko „angielskie". Mocno w duchu tradycji, która nie tylko stworzyła Benthamowski Panoptykon, ale też dalej szuka „klas niebezpiecznych" i domaga się narzędzi wyłuskiwania podejrzanych jednostek na podstawie cech charakterystycznych.

Algorytm

zestaw instrukcji prowadzących do rozwiązania problemu lub wykonania zadania.

Nie mam odruchu obrony swojej ojczyzny i nie będę zaprzeczał. Chciałbym móc powiedzieć, że to wyłącznie nasz problem. Ale tak nie jest. Mówiłem już chociażby o Holandii, ale do listy krajów stosujących podobne narzędzia można dopisać też Hiszpanię. Tam działa między innymi system do oceny ryzyka ekstremizmu. Szybko przekształcił się w narzędzie do śledzenia muzułmanów, bo tak go zaprojektowano. Problem w tym, że niektóre zjawiska, którymi miałaby się zająć sztuczna inteligencja, wcale nie są łatwe do zdefiniowania. A od tego, jakie kryteria „ekstremizmu" czy „dżihadyzmu" przyjmie rząd, zależy, co stanie się potem. Może się stać tak, że masę młodych mężczyzn wrzucimy do bazy danych o ekstremistach, choć nie popełnili żadnego przestępstwa. Nie robią nic zakazanego, lecz zostali zaklasyfikowani tak przez algorytm z powodu swojego wieku, narodowości i religii. Źle zaprojektowany algorytm z miejsca uzna całą kategorię ludzi za podejrzanych – a co potem zrobi z tym państwo, nie wiemy.

Oprócz profilowania pod kątem przestępczości i ekstremizmu warto wspomnieć jeszcze o tym, że w kolejnych krajach Europy sprzedaje się na rynku komercyjnym narzędzia do profilowania osadzonych w zakładach karnych, które obiecują „rzetelną i opartą na danych" ocenę ryzyka recydywizmu. To kolejny ważny obszar.

Ale skoro Unia Europejska ma zakazać profilowania rasowego w algorytmach czy predykcyjnego postępowania policyjnego, to może cała nasza rozmowa nie ma sensu, bo nieetyczne praktyki i tak zostaną objęte zakazami zawartymi choćby w największym rozporządzeniu regulującym użycie sztucznej inteligencji, czyli AI Act?

Bardzo bym się ucieszył, gdyby to była prawda. Oczywiście, Unia Europejska podjęła ważny i odpowiedzialny krok, rozważając zakazanie najbardziej niebezpiecznych technologii wykorzystujących AI. I zgadzam się, że właśnie policyjne algorytmy predykcyjne należy dopisać do tej listy. Co więcej, na pochwałę zasługuje to, że europarlamentarzyści zagłosowali za zakazem. Problem w tym, że zapisy, które ostatecznie trafiły do europejskiego prawodawstwa w AI Act, nie odpowiadają temu, o co wnioskowały organizacje broniące praw człowieka, gdy domagały się zakazu.

Algorytm uczenia maszynowego

(ang. *machine learning algorithm*) – różni się od zwykłego algorytmu tym, że w wyniku jego działania powstają nowe reguły. Taki algorytm nie jest w całości programowany przez człowieka. Wykonuje zadania dzięki analizie przedstawianych danych i obserwacji występujących prawidłowości.

Mówiąc w największym skrócie: zakaz obejmie pewne formy profilowania i zastosowania algorytmów predykcyjnych. Pozostaje jednak szerokie pole do interpretacji, a zarazem otwarta droga do ewentualnego kwestionowania nadużyć w sądach. W rzeczy samej, trochę na to właśnie liczę – że władze będą musiały dowodzić w sądach, że mają prawo do sięgania po tego typu narzędzia. Teraz niestety, w formie rozwodnionej przez Komisję i Radę Europejską, nie możemy mówić o jednoznacznym zakazie. Dopiero się przekonamy, co to będzie oznaczać w praktyce.

I co się wydarzy? Pytam zarówno w krótkim, jak i w dłuższym horyzoncie czasowym.

W XXI wieku, a w ostatniej dekadzie w szczególności, poznaliśmy doskonale tę logikę, która mówi, że trzeba oszczędzać pieniądze i zastępować ludzi maszynami tam, gdzie to tylko możliwe. Technologia będzie tańsza, bardziej efektywna i mniej wymagająca niż ludzie – a przynajmniej tak nam powtarzano. W jakimś sensie teraz to samo, co zrobiono z kasjerami w supermarketach, dzieje się powoli z policjantami i sędziami. Obiecano nam, że technologia rozwiąże wszystkie problemy, o ile my zgodzimy się wszystkie swoje problemy i troski powierzyć w ręce technologii.

Są takie momenty jak protesty Black Lives Matter w 2020 r., gdy cały świat mówi, że nasze instytucje są wadliwe. Drogi wyjścia są dwie. Można naprawić – czy choćby próbować naprawić – złe procedury, porzucić pewne założenia, odrzucić logikę, która normalizuje rasizm. Ale jest też pokusa przeciwna: nałożyć na to wszystko jeszcze narzędzia cyfrowe, algorytmy, sztuczną inteligencję. Wtedy zawsze będzie można powiedzieć, że człowiek nie zawiódł, że w konkretnym rozstrzygnięciu nie ma ludzkiej winy czy rasizmu – takie po prostu są wyroki systemu komputerowego. Jeśli nawet ktoś się pomylił, to na pewno nie policjant, sędzia czy urzędnik – ale komputer.

Łatwo zrozumieć, dlaczego z pewnych względów niektórym taka wizja zautomatyzowania policji i wymiaru sprawiedliwości się podoba. Pozwala rozwodnić odpowiedzialność i w razie niepowodzeń przerzucić winę na technologię. Ale nie musimy się na to godzić.

Jednak czy argument, że zdjęcie obowiązków z barków policji i uwolnienie zasobów wymiaru sprawiedliwości pomoże w skuteczniejszym ściganiu przestępstw, jest bezzasadny?

Generalnie policja nie jest zbyt dobra w ściganiu przestępstw. To, czy w jakiejś dzielnicy pracuje 10 czy 100 policjantów, nie ma przełożenia na to, jaka jest wykrywalność przestępstw. Dużo większe znaczenie ma dostępność miejsc pracy, jakość usług publicznych, poziom zaufania do państwa czy skuteczność walki z biedą i wykluczeniem. Zasoby wydane na poprawę jakości życia mieszkańców, zabezpieczenie społeczne czy edukację mogą się bardziej przyczynić do poprawy bezpieczeństwa. To mój punkt pierwszy. Punkt drugi dotyczy tego, że systemy bazujące na AI też nie przyczyniają się do poprawy bezpieczeństwa. Skuteczność niektórych z nich w ocenie ryzyka nie przekracza 50%, więc nie są lepsze od rzutu monetą. Zazwyczaj modele, na których się opierają, są skrzywione i wskazują, że obcokrajowcy są bardziej skłonni do popełniania przestępstw. Co nijak nie pomaga walczyć z przemocą domową czy przemocą w najbliższej okolicy.

Sądzę więc, że przejdziemy przez podobny cykl jak w przypadku monitoringu w miejscach publicznych w latach 90.

ZAWODNY KOMPAS, CZYLI AI W SŁUŻBIE BEZPIECZEŃSTWU

Sztuczna inteligencja jest dziś wykorzystywana przez policję i inne służby na wiele sposobów. Oto przykładowe zastosowania i kontrowersje, jakie im towarzyszą.

Wykrywanie przestępstw i zakazanych treści w sieci

Algorytmy namierzające treści sprzeczne z prawem czy z tzw. standardami społeczności są w Internecie wszechobecne. Korzystają z nich nie tylko wielkie internetowe platformy – coraz śmielej sięgają po nie także służby. Programy filtrujące spośród wielu innych treści wyławiają te „niechciane", np. pornograficzne, naruszające prawo autorskie czy związane z handlem nielegalnymi substancjami. Dla przykładu popularny Safer pozwala na błyskawiczną analizę materiałów przesyłanych przez użytkowników (zdjęcia, wideo) i namierzenie treści o charakterze seksualnym z udziałem dzieci (CSAM, od ang. *child sexual abuse materials*).

Temat ochrony dzieci budzi szczególne emocje: w Unii Europejskiej powstaje rozporządzenie CSA zwane potocznie kontrolą czatów (Chat Control), które ma nałożyć na operatorów komunikatorów internetowych nakaz filtrowania przesyłanych treści pod kątem CSAM. Niestety projektowana regulacja może nie rozwiązać problemu, a przynieść nowy – w postaci błędów skutkujących oskarżeniami najwyższego kalibru. To nie wszystko: na szali leży los szyfrowanej komunikacji, czyli możliwości komunikowania się w sposób poufny. A ta jest kluczowa nie tylko dla bezpieczeństwa dziennikarzy, adwokatów czy sygnalistów, ale nas wszystkich. Także dzieci.

i w pierwszym dziesięcioleciu xxi wieku. Kiedyś wydawało się, że kamery w miastach uchronią nas przed przestępczością. Dziś jest ich pełno, ale trudno powiedzieć, żeby przyczyniły się do zwiększenia bezpieczeństwa i skuteczności ścigania przestępstw. Mamy więc olbrzymi wzrost nadzoru, ale niekoniecznie jakościową zmianę w sferze bezpieczeństwa.

I tego powinniśmy się obawiać?
Najgorszy scenariusz, który może się zrealizować, jest taki, że te kamery, systemy rozpoznawania twarzy, algorytmy predykcyjne i masowe profilowanie się po prostu przyjmą. I wtedy wszyscy – nie tylko groźni przestępcy – ale wszyscy, którzy podpadli władzom, będą obiektem takiego nadzoru. Uczestnicy protestów ulicznych, członkowie mniejszości narodowych, przeciwnicy władzy... Narzędzia służące rzekomo do walki z przestępczością będą mogły zostać użyte przeciwko nim z wielką łatwością. Nietrudno sobie wyobrazić dzielnice biedy czy imigranckie osiedla zmieniające się w swego rodzaju państwo policyjne napędzane technologiami nadzoru i wolne od niego enklawy zamkniętych osiedli dla bogatych. Podobnie – łatwo sobie wyobrazić skutki trafienia do bazy ekstremistów czy innych „osób wysokiego ryzyka" dla niewinnego człowieka. Gdy pewnego dnia zaczynasz zadawać sobie pytanie „Dlaczego nie mogę dostać pracy, stypendium, miejsca na uczelni czy umówić się do lekarza?" – być może jesteś podejrzany, a nawet o tym nie wiesz.

Wszystko to brzmi jak koszmar, ale tak naprawdę jeśli nie zatrzymamy wdrażania pewnych technologii już teraz, to ta wizja nie jest wcale odległa.

Wykrywanie przestępstw finansowych

Narzędzia bazujące na sztucznej inteligencji wykorzystywane są do typowania nadużyć zarówno w sektorze prywatnym (banki, instytucje finansowe i ubezpieczeniowe), jak i publicznym (administracja skarbowa). Służą one do przeszukiwania dostępnych baz danych (np. transakcji finansowych) w celu odnalezienia wzorców wskazujących na nieprawidłowości. Oflagowanie podejrzanych transakcji może uruchomić dalsze działania, np. zablokowanie konta czy wszczęcie postępowania administracyjnego.

Program Palantir Gotham pokazuje, w jakim kierunku mogą rozwijać się systemy wykrywające nadużycia. Wykorzystuje on uczenie maszynowe do analizy i łączenia danych z różnych, w tym prywatnych, źródeł – m.in. z systemów transakcyjnych, mediów społecznościowych czy metadanych z urządzeń mobilnych. Producent twierdzi, że system jest wielofunkcyjny i może być przystosowany do identyfikowania nie tylko oszustw finansowych, ale też zagrożeń terrorystycznych oraz innych przestępstw.

Polskie państwo nie posunęło się jeszcze tak daleko. Jednak od 2018 r. nasz fiskus także korzysta ze wsparcia zaawansowanych algorytmów. Właśnie wtedy polska Krajowa Administracja Skarbowa (KAS) oficjalnie rozpoczęła korzystanie z Systemu Teleinformatycznego Izby Rozrachunkowej (STIR). Algorytmy STIR są niejawne, więc nie wiadomo, czy wykorzystują techniki uczenia maszynowego. Ale być może to właśnie one przyczyniły się do wzrostu wpływów z podatków w latach 2018–2020.

Dodatkowe środki w budżecie państwa mogą mieć swoją cenę, np. w postaci nadmiarowych kontroli lub oskarżeń kierowanych wobec niewinnych osób. Zaawansowane systemy typujące nieprawidłowości działają zwykle w sposób nieprzejrzysty. Urzędnik nie ma ani czasu na zgłębianie skomplikowanej logiki działania systemu, ani interesu w kwestionowaniu zaproponowanych przez niego rozstrzygnięć. A osobie skrzywdzonej decyzją podjętą na podstawie rekomendacji algorytmu trudno skutecznie się od niej odwołać. Wszystko to paradoksalnie potęguje ryzyko niesprawiedliwych decyzji. A te – jak pokazuje brytyjski skandal pocztowy – mogą mieć dramatyczne konsekwencje (por. *Pomoc społeczna na łasce algorytmów*).

Przewidywanie i wykrywanie zagrożeń w przestrzeni publicznej

Służby korzystają również z systemów predykcyjnych bazujących na szacowaniu ryzyka wystąpienia niebezpiecznych incydentów w konkretnym miejscu i czasie. Działanie takich modeli bazuje na założeniu, że przestępczość nie jest losowa, ale rządzą nią pewne wzorce (godzina, pora roku, liczba skazanych mieszkających w okolicy itp.), które można zidentyfikować i wykorzystać do przewidywania zagrożeń, a co za tym idzie – np. kierowania patroli tam, gdzie będą najbardziej potrzebne. Skuteczność takich programów ma swoje ograniczenia. Dane, które wykorzystują (przede wszystkim policyjne statystyki), często nie oddają aktualnych zagrożeń. Szczególnie trudne do przewidzenia są nietypowe sytuacje. Natomiast przestępcy, którzy wiedzą o sposobie działania policji, mogą modyfikować swoje zachowanie tak, by uniknąć zatrzymania.

Inny sposób funkcjonowania systemów opartych na AI polega na obserwacji otoczenia i próbie wyłapania potencjalnych zagrożeń. Na tej zasadzie działają systemy monitoringu wyposażone w oprogramowanie analizujące obraz z kamer i dźwięk. Wykorzystuje się je np. na dworcach czy na granicy. Wychwytują one m.in. niebezpieczne przedmioty, zbiegowiska, bójki, pozostawianie bagażu, podejrzany sposób poruszania się czy odgłosy wystrzału. W razie wykrycia takiej sytuacji system może np. zwrócić uwagę operatorów monitoringu lub zaalarmować patrol policji. Niestety, zbytnie poleganie na takich alertach może uśpić czujność służb i utrudnić sprawną reakcję w sytuacji, gdy system nie wychwycił zagrożenia. Problemem mogą być także fałszywe alarmy i związane z nimi nadmiarowe kontrole. W praktyce nieproporcjonalnie często dotyczą one grup zmarginalizowanych, np. mniejszości etnicznych, przez co utrwalają systemową dyskryminację.

Ocena ryzyka popełnienia przestępstwa

Jeden z najbardziej kontrowersyjnych i nieprzewidywalnych sposobów wykorzystania sztucznej inteligencji jako wsparcia działań policji i wymiaru sprawiedliwości dotyczy szacowania ryzyka popełnienia przestępstwa przez konkretną osobę. Odbywa się to na podstawie analizy danych zebranych na temat danej osoby i porównania ich z danymi historycznymi dotyczącymi przestępczości. Celnym trafieniom zawsze towarzyszą błędy, które mogą mieć bardzo poważne skutki.

Przykłady takich systemów to holenderski Top 400, czyli lista osób wysokiego ryzyka, które w pierwszej kolejności były poddawane kontroli w przypadku popełnienia przestępstwa przez nieznanego sprawcę, nawet jeśli nic konkretnego nie wskazywało, że miały z nim coś wspólnego, oraz brytyjski system mający wpływ na to, które osoby wchodzące w konflikt z prawem miały postawione w związku z tym zarzuty, a które unikały odpowiedzialności (por. *Cyfrowy policjant nas nie uratuje*).

Duże społeczne oburzenie wywołało działanie COMPAS-u – amerykańskiego systemu szacowania ryzyka recydywy (ocenianego pod kątem wydawania zgody na wcześniejsze zakończenie odbywania kary), a także niestawienia się w sądzie osób oskarżonych o popełnienie przestępstwa. COMPAS zawyżał ryzyko niepożądanych zachowań w stosunku do osób czarnych i kobiet, a jednocześnie zaniżał w stosunku do białych mężczyzn. Podobnie jak pozostałe systemy oceny ryzyka w założeniu miał korygować ludzką uznaniowość w decyzjach podejmowanych – w tym przypadku – przez sędziów, jednak ostatecznie odtworzył systemowe uprzedzenia (zwłaszcza rasowe).

Systemy biometrycznej identyfikacji

Duże kontrowersje budzi również wykorzystywanie sztucznej inteligencji do identyfikacji ludzi. Modele biometryczne są trenowane na danych biologicznych (np. głos, DNA, odciski palca, oczy, ukrwienie skóry, geometria twarzy) lub behawioralnych (np. sposób poruszania się, pisania, czas i wzorce reakcji nerwowych). Biometria stosowana „dla bezpieczeństwa" może polegać na skanowaniu twarzy wszystkich osób pojawiających się w zasięgu kamer monitoringu i porównywania ich z dostępnymi bazami danych. Dane te mogą pochodzić nie tylko z publicznych rejestrów (np. zdjęcia paszportowe), ale także z sieci. Systemy biometryczne mogą działać *ex post* (gdy nagrany materiał służy identyfikacji np. sprawcy przestępstwa) albo – co szczególnie niepokojące – w czasie rzeczywistym (*real time*).

Monitoring wyposażony w rozwiązania bazujące na AI w założeniu miał być wykorzystywany w sytuacjach podwyższonego ryzyka, np. na wielkich imprezach sportowych. Coraz częściej jednak pojawiają się próby wykorzystywania go jako rutynowego narzędzia wspierającego walkę z przestępczością. Budzi to poważne kontrowersje. Na przykład w 2021 r. włoski organ ochrony danych wydał decyzję zakazującą wykorzystywania systemu SARI przetwarzającego dane biometryczne w czasie rzeczywistym. Organ zwrócił uwagę na to, że system naraża wszystkie osoby znajdujące się w jego zasięgu – także te niebędące przedmiotem zainteresowania służb – na ryzyko dyskryminacji.

Działanie takich systemów obarczone jest szeregiem błędów, których skutki mogą być poważne. Oflagowanie przez system oznacza w praktyce aresztowanie i oczekiwanie na wyjaśnienie sprawy za kratami. Błędy oprogramowania wykorzystywanego przez amerykańską policję są już przedmiotem całej serii spraw sądowych. W większości dotyczą one osób o kolorze skóry innym niż biały.

- W 2023 r. Porcha Woodruff – kobieta w zaawansowanej ciąży – została aresztowana i przez 11 godzin była przetrzymywana w celi. Oprogramowanie do rozpoznawania twarzy stosowane przez policję w Detroit zidentyfikowało ją jako podejrzaną o napad i kradzież samochodu. Jak się okazało – niesłusznie. Ostatecznie Woodruff została oczyszczona ze wszystkich zarzutów. Pozwała Departament Policji Detroit o bezprawne aresztowanie.

Potencjał biometrycznej identyfikacji może być wykorzystywany do śledzenia ludzi bez ich wiedzy, zbierania danych o ich życiu prywatnym i osobach, z którymi się kontaktują. W państwach autorytarnych ułatwia prowadzenie systemowego prześladowania grup społecznych lub konkretnych osób. Technologia ta była wykorzystywana przez Chiny np. w stosunku do mniejszości etnicznej Ujgurów oraz do tłumienia protestów w Hongkongu, a przez Rosję – do walki z opozycją i szykanowania protestujących przeciwko wojnie w Ukrainie.

- W 2019 r. Aleksiej Glukhin został aresztowany przez rosyjską policję na podstawie nagrania z moskiewskiego metra. Na nagraniu miał przy sobie kartonową podobiznę aktywisty zatrzymanego za antywojenny protest. Na tej podstawie został oskarżony o udział w nielegalnej demonstracji i ukarany grzywną. Po wyczerpaniu drogi sądowej w Rosji Glukhin złożył skargę do Europejskiego Trybunału Praw Człowieka, który w 2023 r. uznał, że Rosja naruszyła prawo do wolności wypowiedzi oraz prywatności: nie zapewniła mu kontroli nad danymi, które zostały wykorzystane do identyfikacji, ani możliwości zakwestionowania jej wyników.

OPRACOWANIE: Małgorzata Szumańska

WYSSANE Z DANYCH. CO WOLNO SYSTEMOM PROFILUJĄCYM LUDZI

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

AUTORKI:
Anna Obem, Katarzyna Szymielewicz

WSPÓŁPRACA:
Gabriela Bar

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 – por. *AI Act w pigułce*), już teraz przyglądamy się, co nowe prawo może zmienić w relacji między ludźmi a narzędziami wykorzystującymi uczenie maszynowe.

Co to są systemy profilujące i do czego służą

Jeśli czytaliście *Raport mniejszości* Phillipa K. Dicka lub oglądaliście nakręcony na jego podstawie film Stevena Spielberga, to pewnie domyślacie się, czym jest predykcja kryminalna. W dużym uproszczeniu: jest to przewidywanie, kto popełni przestępstwo. W tzw. realu 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" – por. *Cyfrowy policjant nas nie uratuje*).

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 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.

„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.

Podmioty rozwijające systemy należące do kategorii wysokiego ryzyka będą musiały spełnić szereg wymogów. Między innymi:

System zaufania społecznego

(ang. *Social Credit System*, SCS) – system monitorowania i oceny obywateli (a także przedsiębiorstw) działający w Chińskiej Republice Ludowej.

- zarządzić związanym z działaniem tych systemów ryzykiem np. dla 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";
- 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.

Dzięki tego typu procedurom 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".

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.

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.

- Prawo wymaga poinformowania nas o tym, że podlegamy ocenie systemu sztucznej inteligencji.
- System oparty na AI nie może podejmować ważnych decyzji – takich jak przyznanie kredytu, zatrudnienie czy znaczące decyzje medyczne – bez ludzkiego nadzoru. Od tego zakazu są wyjątki, ale nawet w takich przypadkach możemy domagać się, aby ostateczną decyzję podjął człowiek, a nie AI.
- Mamy prawo do wyjaśnienia podstaw rozstrzygnięcia podjętego przy użyciu systemu AI wysokiego ryzyka oraz roli sztucznej inteligencji w procesie podejmowania decyzji.
- Możemy złożyć skargę na źle działający (np. krzywdzący) system AI do organu monitorującego rynek (tj. sprawdzającego, czy systemy AI w Polsce działają zgodnie z prawem).
- W przypadku nieprawidłowości mamy prawo do odszkodowania lub zadośćuczynienia: w obszarze rekrutacji na podstawie 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.

AI ACT W PIGUŁCE

Unijne rozporządzenie w sprawie sztucznej inteligencji, czyli AI Act, to pierwsza próba kompleksowego uregulowania działania AI. Ustanawia ono zasady wprowadzania oraz stosowania systemów sztucznej inteligencji. Dzieli je na cztery kategorie: zakazane, wysokiego ryzyka, ograniczonego ryzyka oraz minimalnego ryzyka.

- **Zakazane** będzie m.in. stosowanie systemów AI pozwalających na ocenę punktową obywateli (tzw. scoring społeczny w stylu chińskiego SCS), przewidujących ryzyko popełnienia przestępstwa przez konkretne osoby lub wykorzystujących słabości danej osoby bądź grupy (związane np. z wiekiem czy niepełnosprawnością).

- Systemy **wysokiego ryzyka** – czyli takie, których stosowanie może wywrzeć szkodliwy wpływ na zdrowie, bezpieczeństwo czy prawa podstawowe (np. systemy decydujące o wynikach rekrutacji czy oceny zdolności kredytowej) – przed wprowadzeniem do obrotu będą musiały przejść specjalne „testy bezpieczeństwa", a ich działanie ma być nadzorowane przez ludzi. Wszystko po to, by zminimalizować ryzyko popełnienia błędu.

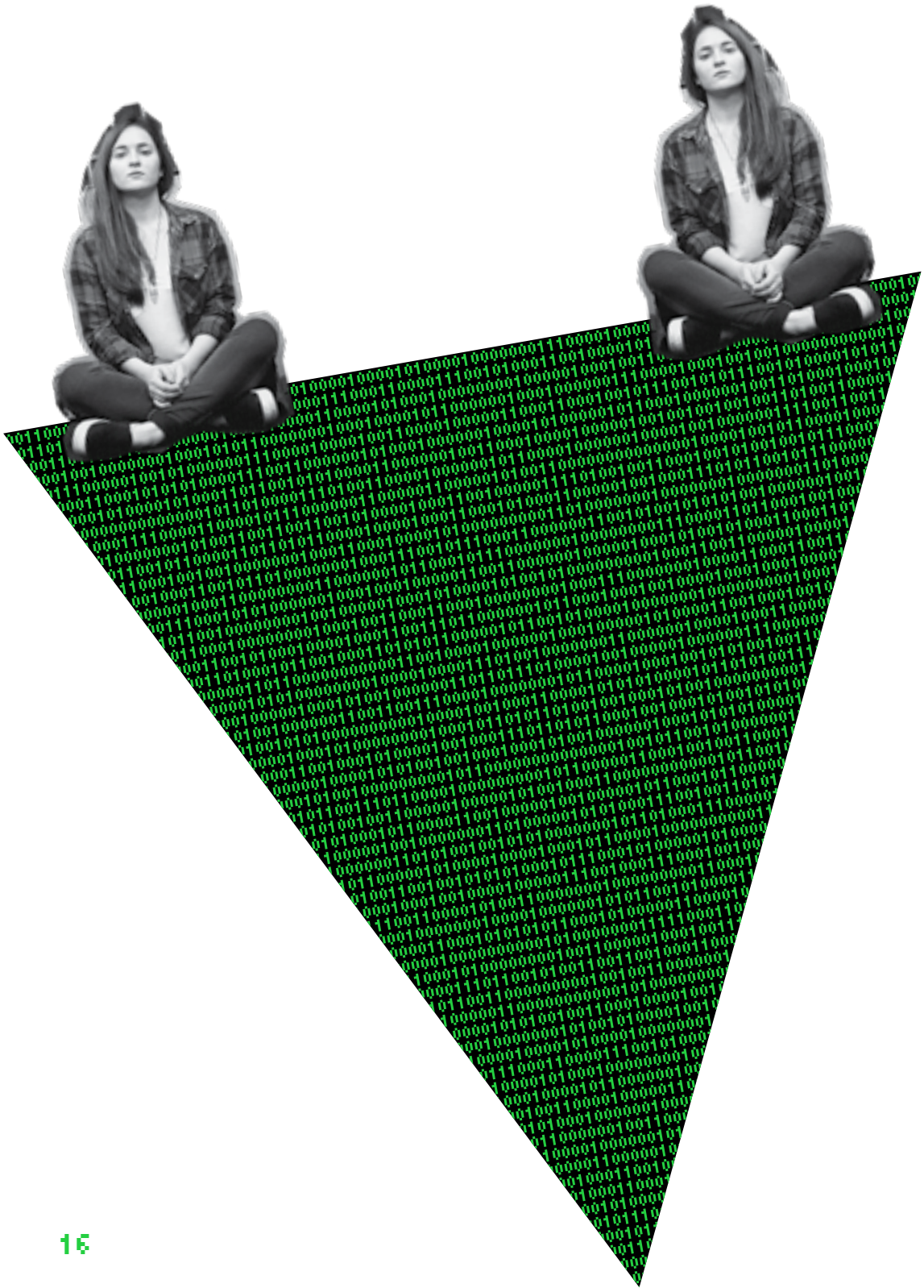
- Kategoria **ograniczonego ryzyka** dotyczy przede wszystkim systemów wchodzących w interakcję z człowiekiem (np. wykorzystujących chatboty) lub tworzących treści (obrazy, teksty lub dźwięki). AI Act wymusza na nich przejrzystość działania, np. poprzez oznaczanie treści wygenerowanych przez sztuczną inteligencję.

- Rozporządzenie daje swobodę korzystania z **pozostałych** systemów (np. takich stosowanych w grach komputerowych).

Wszystkie państwa unijne mają obowiązek powołania organu nadzoru, który będzie kontrolował przestrzeganie obowiązków wynikających z rozporządzenia, a także rozpatrywał skargi na działanie systemów sztucznej inteligencji.

AI Act został przyjęty przez Parlament Europejski 13 marca 2024 r., a wszedł w życie 1 sierpnia 2024 r. Oficjalnie rozporządzenie ma być stosowane od 2 sierpnia 2026 r., jednak część przepisów zacznie działać wcześniej (np. te dotyczące zakazanych praktyk – po 6 miesiącach od wejścia w życie), a niektóre później (dopiero w 2027 r.).

OPRACOWANIE: Mateusz Wrotny



POMOC SPOŁECZNA NA ŁASCE ALGORYTMÓW

Rozmowa Jakuba Dymka z Mikiem Zajko – kanadyskim socjologiem badającym konsekwencje wykorzystywania algorytmów przez administrację publiczną

Jeśli państwo podejmuje decyzję, że komuś należą się albo nie należą się jakieś świadczenia – powinno móc bardzo precyzyjnie wykazać, jak doszło do takiej decyzji i jak ona jest umotywowana. To nie będzie możliwe w przypadku, gdy za proces podejmowania decyzji będzie odpowiadał zaawansowany algorytm filtrujący dane i wyłapujący prawidłowości niewidoczne gołym okiem ani dla przeciętnego człowieka, ani dla doświadczonego urzędnika.

Jakub Dymek: Zajmujesz się kwestią cyfrowego państwa dobrobytu i wykorzystywania algorytmów do nadzoru usług publicznych. Przyznam, że wszystko to nie brzmi specjalnie złowrogo. Może jednak za tymi hasłami kryje się więcej ryzyk, niż się domyślamy?

Mike Zajko: Pojęcie cyfrowego państwa dobrobytu odnosi się do procesu zapoczątkowanego jeszcze w latach 90., w ramach którego do systemów zarządzania świadczeniami społecznymi, pomocą socjalną czy administracją usług publicznych wprowadza się coraz więcej automatycznych mechanizmów decyzyjnych. Z jednej więc strony chodzi o to, by odciążyć człowieka i uprościć biurokrację. Z drugiej zaś – by zmienić rolę ludzi w zarządzaniu świadczeniami, które trafiają do innych, a także naturę pracy w administracji publicznej.

Ryzyko jest takie, że zdejmując ciężar pewnych obowiązków z ludzkich barków, oderwiemy problemy i wyzwania polityki społecznej od kontroli człowieka i uniemożliwimy – niezbędne także w tej dziedzinie funkcjonowania państwa – zrozumienie potrzeb klienta. W konsekwencji narazimy masę niewinnych ludzi na potencjalnie niesprawiedliwe czy niepotrzebnie surowe decyzje podejmowane przez sterowany algorytmem automat.

Komu najbardziej zależało, żeby kwestie pomocy społecznej czy administracji usługami publicznymi zautomatyzować? Rządom i samorządom, które mogą chcieć oszczędzić; firmom, które na wdrażaniu podobnych technologii zarabiają; czy może samym odbiorcom świadczeń i usług…?

W dużej mierze bierze się to z nigdy niekończącej się pogoni za jeszcze większą efektywnością. Oczywiście konkretne motywacje mogą się różnić pomiędzy krajami Europy, Kanadą a USA, ale zbiorczo możemy powiedzieć, że chodzi o wymogi neoliberalizmu, czyli np. o jak najszybsze skrócenie kolejki spraw do rozpatrzenia albo obcięcie wydatków publicznych na urzędników. Impuls przyszedł ze strony rządów – sam model biznesowy dotyczący automatyzowania administracji wykształcił się później.

Kiedy zaczęło się to na dobre?

O cyfryzacji pomocy społecznej pisano już w latach 90., ale zmiana była raczej stopniowa niż nagła. W gruncie rzeczy zaczęło się od przeniesienia do rzeczywistości cyfrowej procesów, które same w sobie były algorytmami – tylko że realizowanymi przez ludzi. Wszystkie te formularze, przetwarzanie informacji, nanoszenie ich

na arkusze i wyciąganie z nich systemowych wniosków to także były algorytmy, ale wykonywane przez człowieka. Istnieją badacze, którzy wskazują, że państwo tak chętnie adaptuje się do nowych rozwiązań zmierzających do automatyzacji, bo samo jest czymś na kształt systemu przetwarzania informacji i wdrażania decyzji, czyli – jak powiedzieliby niektórzy – formą sztucznej inteligencji. Pokazywanie związków i analogii między algorytmami a funkcjonowaniem administracji publicznej jest dziś szczególnie zasadne.

Dlaczego?

Bo urzędy czy organy administracji publicznej coraz mocniej wdrażają systemy cyfrowe oparte na algorytmach, a dzięki temu my coraz wyraźniej widzimy, że działanie aparatu państwa było algorytmiczne na długo przed naszą cyfrową teraźniejszością. Wejście algorytmów cyfrowych zaczęło się od arkuszy w Excelu. Potem nastąpiła dalsza automatyzacja procesów, która zaczęła przypominać coś na kształt AI, zaś ostatnia dekada (a właściwie ostatnie pięć lat) to wdrożenie systemów opartych na uczeniu się maszynowym, czyli właściwej sztucznej inteligencji – tak jak rozumiemy ją dzisiaj.

„Właściwej", czyli jakiej?

Fachowo mówi się o zautomatyzowanych systemach decyzyjnych (*automated decision-making systems*, ADM) w dziedzinie administracji publicznej. I rzeczywiście jest to bardziej odpowiedni termin. Powinniśmy przestać kojarzyć AI z jakimiś myślącymi maszynami czy robotami wyposażonymi w mózgi i dysponującymi „inteligencją".

Powinniśmy unikać ekscytacji sztuczną inteligencją w administracji publicznej, bo ten termin utrudnia nam zadawanie właściwych pytań. Choćby dotyczących relacji między zautomatyzowanymi systemami, decyzjami ludzi (urzędników publicznych) i aparatem biurokracji.

Prawdziwie autonomiczny system podejmowałby decyzje na temat jednostek na podstawie szerokiego zbioru danych. Decyzje te miałyby moc wiążącą i były powszechnie uznawane. Tak jeszcze nie jest.

Jednak te systemy są „inteligentne". To nie jest prosta automatyzacja, gdzie kod komputerowy przejmuje pewne rutynowe czynności od człowieka w co najmniej jeden istotny sposób. W zastosowaniu ADM w cyfrowym państwie dobrobytu chodzi nie tylko o to, by algorytm sprawdził, czy człowiek właściwie wypełnił wniosek. Algorytm jest trenowany na wielkich zasobach danych w procesie uczenia maszynowego, aby był zdolny rozpoznawać prawidłowości i odpowiednio je klasyfikować w kolejnych zbiorach danych. Dzięki temu narzędzia są bardziej sprawne, ale i nieprzejrzyste – patrząc z zewnątrz, nie jest do końca jasne, w jaki sposób dochodzą do podejmowanych decyzji.

Ale o jakim stopniu skomplikowania tak naprawdę mówimy? Jestem w stanie wyobrazić sobie zarówno bardzo banalne, jak i bardzo wyrafinowane zastosowania podobnych narzędzi.

Przykładowo „tropienie wyłudzania zasiłków". Gdy używa się takiego sformułowania, natychmiast pojawia się podejrzenie, że chodzi o masowe zjawisko i działalność kryminalną. W rzeczywistości „wyłudzanie" może być wynikiem zmieniających się przepisów, błędów w danych czy zmieniających się kryteriów. Ale to jest ciekawe nawet na tym poziomie. Zauważmy: już u podstaw leży założenie, że ludzi pobierających zasiłki z całą pewnością jest za dużo i że na pewno mamy do czynienia z nadużywaniem systemu.

W tym kontekście wyzwaniem pozostaje tylko to, jak wyeliminować nadużycie. Algorytmy przygotowane do przesiewania i porównywania dużych zasobów danych i weryfikowania odrębnych baz danych mogą być w tym pomocne. Spójrzmy na Holandię, gdzie m.in. na podstawie pomiarów zużycia wody i liczby osób zameldowanych pod konkretnym adresem szukano „czerwonych flag" – czyli potencjalnie podejrzanych zachowań do sprawdzenia. Najczęściej system nie działa w pełni automatycznie, tylko referuje oflagowany adres do człowieka.

Czyli decyzja, co z tym dalej zrobić, nadal leży po stronie urzędnika?

Tak, bo mówimy w rzeczywistości o dwóch rzeczach, które można zautomatyzować: o mechanizmie coraz bardziej autonomicznego wyszukiwania i łączenia różnych zbiorów danych oraz o samym mechanizmie decyzji

administracyjnych. Na razie wygodniej jest zrobić to pierwsze. Nie oznacza to jednak, że decyzja człowieka nie jest w pewnym wymiarze podyktowana przez działanie algorytmu – przecież sam zasób „podejrzanych spraw" i ścieżka prowadząca do odnajdywania „czerwonych flag" zostały już delegowane na algorytm.

W Holandii doprowadziło to do słynnej afery zasiłkowej, w której główną rolę odegrał system SyRI. Przez kilka lat tysiącom rodzin kazano zwracać zasiłki społeczne, które rzekomo wyłudzili. W 2019 r. ujawniono, że w rzeczywistości za decyzjami stoi nieznany wcześniej publicznie algorytm. System ten, posługując się uczeniem maszynowym właśnie (m.in wspomnianymi danymi o zużyciu wody), wyszukiwał „podejrzanych". Holenderski Urząd Podatkowy, który zarządzał całym procesem, robił to dodatkowo w sposób nieprzejrzysty, a logika podejmowania decyzji była sprzeczna – jak wskazywały różne urzędy – z prawem i ogólnie przyjętymi w Holandii standardami. Pojawiły się kary i oskarżenia o łamanie prawa do prywatności. Ostatecznie wskutek skandalu rząd Marka Ruttego podał się w 2021 r. do dymisji [później ponownie wygrał wybory – przyp. JD].

Czy to jedyna taka historia?

Nie. W Wielkiej Brytanii miał miejsce tzw. skandal Horizon, który korzeniami sięga jeszcze lat 90. Chodzi o system, który miał nadzorować funkcjonowanie placówek pocztowych, prowadzonych przez jednoosobowe działalności gospodarcze na podstawie kontraktu z brytyjską pocztą. System Horizon wyprodukowany przez firmę Fujitsu miał prowadzić księgowość i inwentarz małych placówek pocztowych. W wyniku błędów zaczął pokazywać braki w kasie albo zaopatrzeniu. Pocztowcy zgłaszali błędy, ale poczta utrzymywała, że z systemem wszystko jest w porządku, domagając się zwrotu brakujących kwot z kieszeni tych ludzi. To trwało latami. Reklamacje i odwołania były ignorowane, w wielu przypadkach sprawy trafiały do sądu, ludzi skazywano na więzienie, tracili oszczędności, niektórzy odebrali sobie życie. Ostatecznie rząd i zarząd poczty musieli przyznać się do błędu, a kwoty przeznaczone na odszkodowania i zadośćuczynienia idą w setki milionów funtów.

W Australii podobny skandal dotyczył systemu automatycznego naliczania należności i długów od obywateli, który od samego początku działania w 2016 r. budził wątpliwości. System funkcjonował zaledwie kilka lat. Już w 2020 r. rząd obiecał się z niego wycofać, a wadliwie naliczone długi umorzono w 500 tys. przypadków. Coś podobnego przetestowano w stanie Michigan w USA – z podobnym skutkiem. Wszystkie te systemy wcale nie musiały być bardzo skomplikowane, żeby dochodziło do błędów. Przeciwnie – wystarczyło, że kwoty, na podstawie których system dokonywał obliczeń, były nieprawidłowe.

Na tych przykładach widać, że rozwiązania, które miały przynieść oszczędności i zwiększyć efektywność działania, doprowadziły ostatecznie do konieczności wypłaty przez rząd wielkich odszkodowań i ponoszenia znacznie większych kosztów.

Ktoś jednak powie, że o ile proste systemy skazane są na głupie błędy, to wejście tych algorytmów na wyższy poziom zaawansowanie pozwoli je wyeliminować. I że to kwestia techniczna – bardziej precyzyjne i operujące na wielkim poziomie szczegółowości narzędzia nie będą powielać tak banalnych pomyłek.

Myślę, że bardziej skomplikowane systemy, właśnie przez stopień zaawansowania, wprowadzają nowe punkty węzłowe, gdzie więcej spraw może pójść nie tak. Jedną ze znanych prawidłowości związanych z AI jest to, że im bardziej efektywna jest sztuczna inteligencja i uczenie maszynowe, tym mniej przejrzyste i trudniejsze do wytłumaczenia jest ich działanie. A to jeden z podstawowych wymogów systemów, które używane są w administracji publicznej. Tym bardziej w obszarach państwa, które dotyczą równego traktowania czy wymiaru sprawiedliwości.

Jeśli państwo podejmuje decyzję, że komuś należą się, albo nie należą się jakieś świadczenia – powinno móc bardzo precyzyjnie wykazać, jak doszło do takiej decyzji i jak ona jest umotywowana. To nie będzie możliwe w przypadku, gdy za proces podejmowania decyzji będzie odpowiadał zaawansowany algorytm filtrujący dane i wyłapujący prawidłowości niewidoczne gołym okiem ani dla przeciętnego człowieka, ani dla doświadczonego urzędnika. Żeby ten system był zrozumiały, należałoby go jakoś uprościć.

Ale nawet uproszczony – pozostanie czarną skrzynką. To znaczy, że choć byśmy wiedzieli w zarysach, co i jak działa, to mechanizmy pozostaną ukryte i niemożliwe do rozgryzienia.

A gdyby rządy ujawniały więcej informacji o funkcjonowaniu tych systemów?

Główny problem z tym postulatem polega na tym, że udostępniane informacje i tak nie pozwolą się rozeznać w działaniu konkretnych narzędzi i zrozumieć, jaka relacja człowiek–maszyna naprawdę zachodzi. Rządowe agencje twierdzą też, że w ich ocenie żadne decyzje nie są podejmowane przez ADM-y faktycznie automatycznie, a ostatecznie decyduje o wszystkim człowiek. Jednak zgadzam się co do tego, że sama wiedza o wykorzystywaniu podobnych systemów może być pomocna. Dotychczas dogadywaliśmy się o wielu przykładach po fakcie. Tak było w Kanadzie w przypadku używania ADM-ów do zarządzania migracją.

Na czym to polegało?

W Kanadzie była długa kolejka nierozpatrzonych wniosków wizowych i deficyt urzędników zdolnych procesować na czas kolejne wnioski. Wprowadzono więc automatyczny system, na początku ograniczony m.in. do Indii i Chin, skąd spływało bardzo dużo wniosków. System filtrował je, z jednej strony dając zielone światło obiecującym kandydatom, a z drugiej – szukając schematów, które pozwoliłyby wskazać ewentualne oszustwa. Założenie było takie, że system nie będzie odrzucał wniosków sam z siebie – to zawsze miał robić człowiek. W teorii nie mogło się zdarzyć, że komuś odmawia się prawa wjazdu do Kanady na podstawie decyzji automatu. Urzędnicy dostawali tylko pewne komunikaty, które miały ich przygotować do podjęcia decyzji.

W urzędzie działał jednak także system nazwany Chinook, który pozwalał udzielać odpowiedzi na zasadzie „kopiuj–wklej". Na tej zasadzie formułowano listy zawierające odmowę przyznania wizy. Ich treść nie pozwala już mówić o tym, że dochodziło wnikliwej i samodzielnej oceny przypadków przez urzędników, którzy w teorii mieli rozpatrywać te sprawy indywidualnie. Prawnicy reprezentujący migrantów odkryli, że fakty w listach przestały pasować do konkretnych osób i ich sytuacji. Organizacje pozarządowe czasami wyciągały taką sprawę na jaw po latach... Jest to jeden z tych obszarów działania algorytmów, którym społeczeństwo niespecjalnie się interesuje. Znaczna część procesu automatyzacji dokonuje się za kurtyną niewiedzy.

Bo dotyczy osób biedniejszych i ludzi z dala od szczytów społecznej drabiny?

Systemy nadzoru zazwyczaj służą najbogatszym. Są wymierzone przeciwko grupom słabszym – biedniejszym, marginalizowanym, migrantom – poddanym już różnym reżimom kontroli społecznej. Dzieje się tak nawet w przypadku systemów wykrywania oszustw podatkowych czy unikania opodatkowania, które teoretycznie powinny uderzać w najbogatszych.

> Automatyzacja czy wprowadzenie algorytmów nie sprawi, że dochodzenie zaległości podatkowych od wielkich korporacji czy bogaczy, których stać na doskonałych prawników, będzie dla administracji państwowej tańsze. Osoby decyzyjne w organach skarbowych mogą chcieć od tego odstąpić, bo mają świadomość, że najzamożniejsi będą się długo i skutecznie bronić – a zamiast tego rodzi się pokusa, żeby używać tych systemów wobec ludzi, którzy takiej możliwości nie mają.

Czyli koniec końców w cyfrowym państwie dobrobytu zawsze pojawia się konflikt wartości – choćby między efektywnością a zasadami demokratycznymi?

Myślę, że widzimy napięcie między technokracją (imperatywem efektywności i cięcia kosztów) a demokracją z jej imperatywami praw człowieka (rozumianymi także jako prawo do prywatności i kontroli nad decyzjami, które nas dotyczą). I widzę, że ciążą one w przeciwstawnych kierunkach.

Czarna skrzynka

system, którego wewnętrzne zasady działania są niejasne lub niezrozumiałe dla człowieka.

PRAWDA CZY MIT?
Wybierz prawdziwe zdanie.
Tylko jedna odpowiedź jest prawidłowa.

- [ ] AI to humanoidalny robot
- [ ] AI zna odpowiedź na każde pytanie
- [ ] AI mówi prawdę
- [ ] AI przewiduje przyszłość
- [ ] AI się nie myli
- [ ] AI jest obiektywne
- [ ] AI rozwiąże problemy ludzkości
- [ ] AI świadomie nas unicestwi
- [x] Panoptykon rozbraja mity dotyczące AI*

WSPIERAJ PANOPTYKON! REGULARNIE.

Nr konta:
43 1440 1101 0000 0000 1044 6058
panoptykon.org/wspieraj

* Na stronie panoptykon.org znajdziesz rzetelne informacje o AI. Za darmo.

CZEMU SŁUŻY SZTUCZNA INTELIGENCJA

Interesom firm z Doliny Krzemowej? Cięciu kosztów? Osłabieniu pozycji pracowników? Zdjęciu odpowiedzialności z osób podejmujących trudne decyzje? Ochronie naszego wygodnego świata przed czyhającym za granicą „innym”? Profilowaniu ludzi na masową skalę? A co z interesem społecznym i dobrem wspólnym? Prezentujemy fragmenty wywiadów, które rzucają światło na to, co decyduje o kierunku rozwoju i sposobach wykorzystania sztucznej inteligencji. Pełne wersje wywiadów – zrealizowanych przez Jakuba Dymka – dostępne są na stronie panoptykon.org.

CATHY O'NEIL:

Bloomberg zapytał sztuczną inteligencję o to, jak wygląda nauczyciel. Otrzymał dużo więcej wizerunków kobiet, niż wynikałoby to z faktycznych proporcji płci w zawodzie. Ktoś powie, że nieudane, seksistowskie czy dziwaczne obrazki to żaden problem. OK, zgoda. Ale mnie nie interesują obrazki generowane przez AI tylko fakt, że ten sam system – bez żadnej sensownej weryfikacji – zaraz będzie używany w rekrutacji, w ochronie zdrowia i we wszystkich innych dziedzinach, gdzie nawet drobne różnice będą miały poważne konsekwencje.

Ludzie, z którymi rozmawiam, tłumaczą, że prędzej czy później argument oszczędności czy optymalizacji kosztów zawsze wygrywa. Ktoś przychodzi do agencji rządowej, ministerstwa czy jakiejś instytucji samorządu i mówi: „Dlaczego tej żmudnej i mozolnej pracy nie robi jeszcze komputer?".

Oczywiście, algorytmy czy sztuczna inteligencja sprawdzają się doskonale w jednym zadaniu, czyli krótkoterminowym rozwiązywaniu problemów budżetowych narzucanych przez neoliberalizm i politykę oszczędności. Presja na to, by ciąć budżety, narasta – aż dochodzimy do momentu, w którym uznajemy ludzi za zbędny koszt. Problem w tym, że tzw. komputer często nie wykonuje tej pracy, w której miał ludzi zastąpić.

Efektywność to nie wszystko. Mamy do czynienia z czymś znacznie poważniejszym niż pytanie o to, czy komputer działa, czy nie, albo czy konkretne cięcia budżetowe wdrażane przez rządy są uzasadnione.

To znaczy?

Mówimy o problemie przepływów w całej globalnej gospodarce. Ludzie tracą pracę w dziedzinie, w której byli wyspecjalizowani. A wraz z likwidowaniem ich miejsc pracy znika zasób doświadczenia i poziom zawodowego rzemiosła, które reprezentowali, prawda? Kto na tym zyskuje? Ci, którzy zastępują to skomercjalizowanym, fatalnym, wadliwym modelem zastępczym. Dolina Krzemowa. To jest niesłychanie niesprawiedliwe i na dłuższą metę niewydolne.

Chcesz powiedzieć, że cięcia budżetowe dokonywane przez rządy i zastąpienie „drogich" miejsc pracy ludzi „tanimi" algorytmami jest tak naprawdę formą transferu finansowego z państwowej kasy na konta firm z Doliny Krzemowej?

Dokładnie tak. Na 100%. To jest mówienie ludziom, którzy dzięki sektorowi publicznemu mieli normalną pracę, godne życie i zabezpieczoną emeryturę: „Dziękujemy, już was nie potrzebujemy. Pieniądze na wasze pensje lepiej wydamy w Dolinie Krzemowej".

Co należałoby w tej sytuacji zrobić?

Zamiast dyskutować o naukowym aspekcie danych, o matematyce, o zawiłościach kodu komputerowego – powinniśmy rozmawiać o uczciwości rozwiązań. Ktoś może powiedzieć: „Technologia to nie dla mnie, nie znam się", ale czy może powiedzieć: „Uczciwość mnie nie obchodzi"? 95% dyskusji o zastosowaniu algorytmów w biznesie czy polityce publicznej można zastąpić dyskusją o uczciwości i o ważeniu interesów różnych zainteresowanych grup.

Nie pomaga mit geniusza, który otacza przedsiębiorców z Doliny Krzemowej. Twórcy AI przedstawiają się jak bogowie. Mówią nam, że tworzą maszynę obdarzoną świadomością. Jeśli uwierzymy, że tylko ktoś nadzwyczajny jest w stanie zrozumieć AI, nie będziemy zdolni jej kwestionować. Osobiście proponuję podejście dokładnie przeciwne: przyglądajmy się sztucznej inteligencji krytycznie. Każdy ma prawo zapytać: „Czy ktoś właśnie aby nie chce mnie oszukać"?

CHRIS JONES:

Unia Europejska pracuje nad systemem pozwoleń na przemieszczanie się nazywanym European Travel Information Authorization System. W jego ramach dane od podróżnych, nawet tych, którzy nie potrzebują wiz, będą przetwarzane przez algorytm. Na jakich podstawach i na bazie jakich kryteriów algorytm będzie podejmował różne decyzje – to oczywiście niejawne. Przetwarzane przez komputery dane milionów ludzi będą wykorzystywane do opracowywania jeszcze bardziej rozwiniętych procedur i mechanizmów przetwarzania kolejnych danych – i tworzenia jeszcze dalej idących algorytmów. Mówimy o profilowaniu ludzi na masową skalę, ale także na indywidualnym poziomie, ponieważ każdemu przypisany zostanie „profil ryzyka".

I czym to skutkuje?

To niekoniecznie oznacza, że osobie profilowanej pod tym kątem odmówi się wjazdu albo zakaże przekraczania granic. Ale może wpadniesz w jakąś szufladkę? Może powinniśmy powiadomić jakieś służby w twojej ojczyźnie albo w innym państwie członkowskim o tym, gdzie jesteś? Może twoje dokumenty powinny zostać prześwietlone bardziej dokładnie?

Mówimy w gruncie rzeczy o tym, że niesprawdzona jeszcze technologia zostanie przetestowana na setkach milionów ludzi. A nie mamy żadnych dowodów, że coś podobnego jest potrzebne. Nie widziałem żadnych badań, które wskazywałyby, że miliony osób podróżujących do Europy każdego roku trzeba objąć aż takim nadzorem, bo stanowią zagrożenie. A ten program obejmie także w przyszłości osoby, które już mają wizy – i będzie się dalej rozszerzał.

Czy to będzie zautomatyzowane?

W teorii nie ma prawnej możliwości, żeby proces decydowania o zezwoleniu lub odmowie wjazdu realizował sam algorytm bez odpowiedzialności człowieka. Prawo na poziomie unijnym tego zabrania. Jednak choć nie ma żadnych planów, aby podobny proces zautomatyzować w najbliższej przyszłości, nie oznacza to, że decyzje algorytmów nie będą miały wpływu na decyzje urzędników.

Dlaczego?

Podobne systemy dostarczają urzędnikom gotowych rekomendacji czy ocen, które w oczywisty sposób wpływają na ich późniejsze kroki.

Można więc powiedzieć, że i tu dochodzi do pewnego rodzaju oddalenia czy wypchnięcia trudnych albo kontrowersyjnych decyzji i przeniesienia odpowiedzialności za nie na rozwiązanie cyfrowe.

To zjawisko, które nazywamy outsourcingiem moralnym. Bazuje ono właśnie na tym, żeby decyzje o ciężarze moralnym – które mogą kogoś potencjalnie skrzywdzić albo prowadzić do wyboru mniejszego lub większego zła – odsunąć od ludzi je podejmujących, a przenieść np. na algorytmy. Możemy sobie więc wyobrazić przyszłość, w której ten moralny outsourcing zostaje naprawdę wdrożony. I system komputerowy ocenia, komu bardziej należy się ochrona międzynarodowa czy zezwolenie na wjazd: osobie chorej, cierpiącej z powodu zaburzeń psychicznych, doświadczającej przemocy domowej, wywodzącej się z mniejszości narodowej, a może jeszcze komuś innemu? O ile łatwo sobie wyobrazić sam proces, trudniej założyć, że algorytm zdolny byłby do uczciwej oceny. Chciałbym jednak zaznaczyć, że choć to jest idea, o której się w unijnych instytucjach rozmawia, nie mamy pewności, czy rzeczywiście wdrażanie jej w życie idzie do przodu. Ale czy coś podobnego można już testować? Jak najbardziej.

Jak może zatem wyglądać przyszłość praw człowieka i ochrony europejskich granic w miarę postępów technologii?

Ironia polega na tym, że nie musimy wybiegać wyobraźnią daleko w przyszłość. Wystarczy zauważyć, co dzieje się wokół nas. Miliony ludzi codziennie poruszają się po Unii Europejskiej i przekraczają jej granice całkowicie normalnie – choć za cenę tego, że zrzekają się części swojej prywatności (muszą się zgodzić na skany tęczówki czy odcisków palców). Ale oprócz nich są inni. Ci, którzy urodzili się w niewłaściwym kraju, w złym miejscu i czasie – oni nie mogą cieszyć się luksusem swobodnego poruszania po UE. Nieważne, czy szukają schronienia, czy czegokolwiek innego – muszą znaleźć inne sposoby, by tutaj się dostać. Wszystkie rozwiązania prawne czy technologiczne, które dziś mamy, służą tylko podtrzymaniu i wzmocnieniu tej nierównowagi. A więc problemy leżą gdzie indziej niż rozwiązania, które miałyby im rzekomo zaradzić.

PARIS MARX:

Dolina Krzemowa przez długi czas przekonywała nas, że stanowi ucieleśnienie dobrego kapitalizmu: wszystkie te slogany typu „Don't be evil", owocowe czwartki i luźne podejście do hierarchii w pracy. Gdy jednak spojrzeć na to, co robi naprawdę, staje się jasne, że chodzi o osłabienie pracowników, a nie o żaden lepszy kapitalizm. Żeby to w pełni wyjaśnić, musimy się cofnąć do poprzedniej dekady.

OK. Dlaczego?

Oczywiście nowe technologie wielokrotnie w przeszłości były wykorzystywane do tego, aby osłabić pozycję przetargową pracowników. Ale po 2010 r. byliśmy świadkami narracji na temat tego, że autonomiczne samochody mają pozbawić pracy taksówkarzy, dostawców, kierowców ciężarówek. Sporo mówiło się zatem, co należy zrobić, aby zapewnić tym ludziom byt. Może wprowadzić gwarantowany dochód podstawowy? A jednak zaobserwowaliśmy zupełnie inne zjawisko.

Miejsca pracy nie zniknęły?

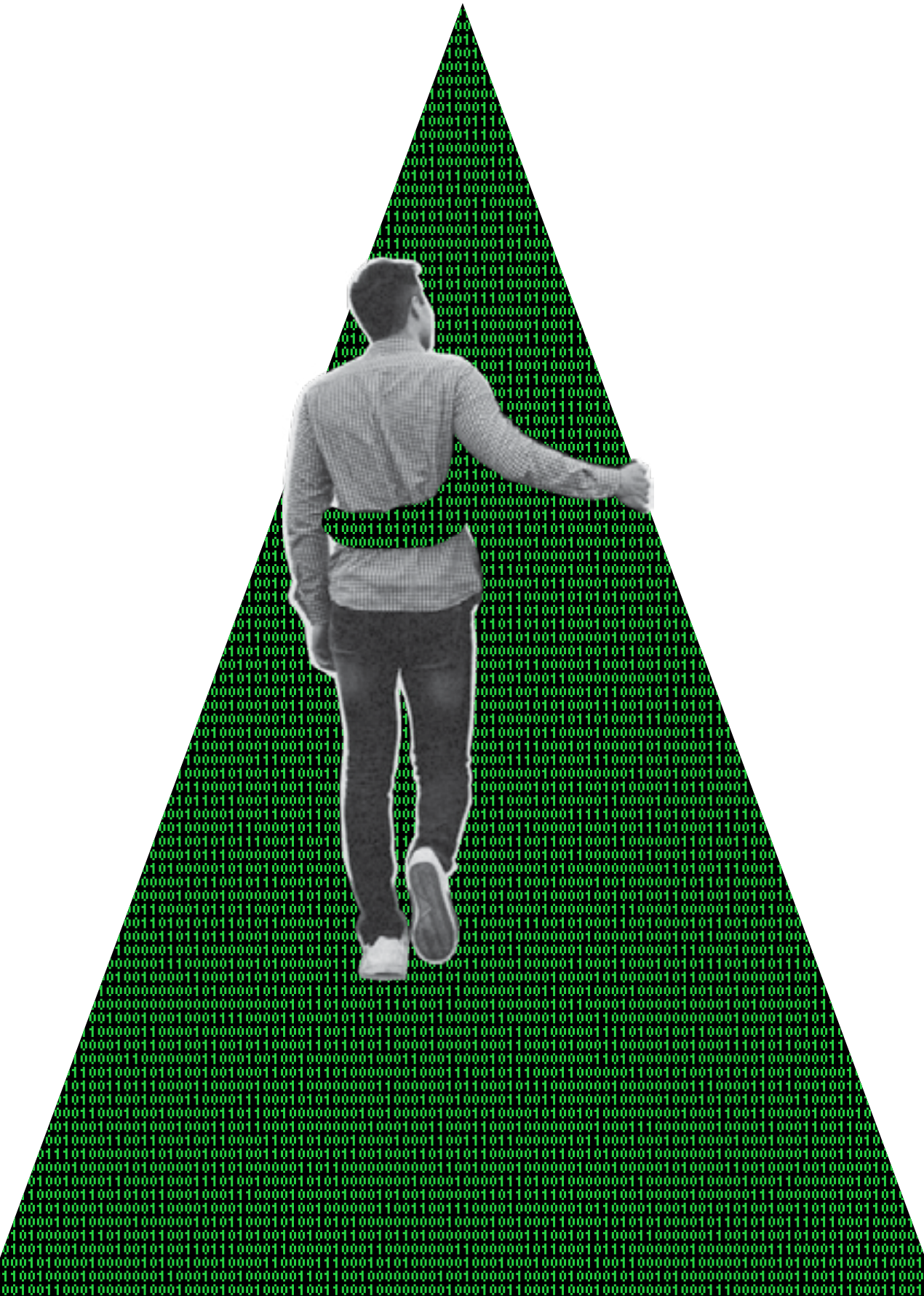
Nie. Firmy użyły nowych technologii (algorytmów, modeli predykcyjnych, danych) nie do pozbawienia ludzi pracy, lecz do przekształcenia relacji pracownik–platforma i do zmiany układu sił na swoją korzyść. Dużo więcej miejsc pracy poddano tzw. uberyzacji: wyprowadzono je z sektorów dobrze płatnych i uzwiązkowionych do zautomatyzowanych centrów logistycznych, gdzie zatrudnia się ludzi przez agencje pracy tymczasowej i stosuje umowy czasowe. W handlu coraz więcej pracy dla wielkich cyfrowych gigantów zlecano podwykonawcom. Cybergiganci nie mogli zakazać działalności związkom zawodowym u swoich dostawców, ale mogli nie przedłużać kontraktów tym firmom, gdzie zawiązały się organizacje pracownicze – dzięki czemu osiągnęły podobny efekt. Oczywiście algorytmy zostały wprowadzone także do zarządzania i nadzoru pracowników w czasie rzeczywistym oraz do śrubowania i egzekwowania norm.

Czyli w tym całym haśle, że „AI zabierze ludziom pracę", nie chodzi o to, że praca dosłownie zniknie – ale że cały technologiczny progres posłuży temu, aby pogorszyć sytuację pracownika, a algorytmy wykorzystać do dyscyplinowania i nadzoru?

Dokładnie. Nie chodzi jednak o technologię jako taką, lecz o sposób, w jaki używana jest ona przeciwko pracownikom. To bardzo stare zjawisko. Nowe jest tylko to, że dziś mówimy o algorytmach czy sztucznej inteligencji. Technologie cyfrowe są przedłużeniem czegoś, co znamy od bardzo dawna: obniżania siły przetargowej pracowników za pomocą groźby automatyzacji i likwidacji ich miejsca pracy. Gdy mówimy, że „AI zabierze ludziom pracę", wydawać się może, że jakaś nieokreślona siła zrobi ludziom krzywdę. Jednak nie chodzi o to, co jakaś technologia zrobi sama z siebie. Miejsca pracy są zagrożone ze względu na decyzje, jakie podejmuje się na najwyższym piętrze w gabinetach prezesów. Widzimy, jak łatwo widmo technologii może zostać użyte jako narzędzie nacisku: „Musicie zaakceptować dzisiejsze warunki pracy, bo jutro i tak może was zastąpić technologia". Opowiadanie, że ludzie stracą pracę w wyniku jej rozwoju, jest użytecznym sposobem na przekierowanie strachu (na jakieś roboty i sztuczną inteligencję) – tak by firmy mogły dalej robić to, co już robią.

Don't be evil

(Nie bądź zły) – dawne motto firmy Google. Po restrukturyzacji w ramach holdingu Alphabet w 2015 r. zmieniono je na: „Do the right thing" („Postępuj właściwie").



CZY TE BOTY MOGĄ KŁAMAĆ

Wirtualny asystent linii lotniczej podał klientowi błędne informacje. Firma próbowała wykręcić się od odpowiedzialności, ale sąd uznał, że powinna zrealizować obietnicę złożoną przez bota funkcjonującego na jej stronie internetowej. Co mówi AI Act o odpowiedzialności za słowa i działania botów?

AUTORKI:
Anna Obem
Katarzyna Szymielewicz

WSPÓŁPRACA:
Gabriela Bar

Czy Alexa, Max i Siri podejmą za ciebie decyzję

Jeszcze kilka lat temu rewolucją w obsłudze klientów były chatboty oparte na założonych z góry scenariuszach rozmów. Dzisiaj wirtualni asystenci to nie tylko coraz bardziej zaawansowane czaty. To także asystenci w rodzaju Alexy czy Siri, które mogą obsługiwać konkretne funkcje smartfona, albo programy automatyzujące pracę w stylu Microsoft Copilot.

Niektóre takie narzędzia, jak bot Max firmy Orange, otwarcie działają w interesie jednej firmy (np. rozwiązują konkretne sprawy). Inne, jak Microsoft Copilot, są reklamowane jako „towarzysze internetowych peregrynacji". Zrobią za ciebie nużący risercz i jednoznacznie odpowiedzą na pytanie. To kolejne wcielenie internetowej wyszukiwarki, ze wszystkimi jej zaletami i wadami, tylko spotęgowane. Narzędzia, które selekcjonują informacje, mogą przecież kształtować decyzje konsumenckie. A czasami podpowiadać produkty droższe, tańsze albo po prostu należące do wspierającej narzędzie firmy.

Marzenia o inteligentnych asystentach nie są obce administracji, również polskiej. W wywiadzie dla Dziennika Gazety Prawnej dyrektor Centralnego Ośrodka Informatyki Radosław Maćkiewicz zapowiedział, że taki asystent ma zacząć działać na stronach gov.pl.

„Zaczęliśmy prace nad funkcjonalnością chatbota wykorzystującego sztuczną inteligencję. Najpierw chcemy uruchomić go na stronie gov.pl, żeby pomógł nam załatwiać sprawy urzędowe. Dzięki niemu, zamiast przeglądać portal, będzie można zadać pytanie w języku naturalnym i uzyskać odpowiedź (...). Wykorzystujemy model od OpenAI, choć zostanie on ograniczony do informacji, które są na gov.pl, żeby nie odpowiadał w dziwny sposób".

Chatboty: co zmieni AI Act

„Dzień dobry, jestem Max, twój wirtualny asystent" – takie powitania to już norma na wielu infoliniach. AI Act stawia kropkę nad i. Według art. 50 ust. 1 dostawcy systemów AI „wchodzących w bezpośrednią interakcję z osobą fizyczną" mają obowiązek poinformowania tej osoby o tym, że ma do czynienia z systemem AI (a nie człowiekiem, np. pracownikiem firmy).

Zostaniemy też poinformowani, że rozmawiamy z chatbotem, jeśli ten będzie „służył ogółowi społeczeństwa na potrzeby składania zawiadomień o popełnieniu przestępstwa".

Kiedy nie dowiemy się, że rozmawiamy z botem?

Chatbot

program komputerowy, którego zadaniem jest prowadzenie konwersacji z człowiekiem.

Po pierwsze, jeśli jest to „oczywiste" – z punktu widzenia „osoby fizycznej, która jest dobrze poinformowana, spostrzegawcza i uważna, biorąc pod uwagę okoliczności i kontekst użycia". Taką „oczywistością" będzie zapewne bot w czasie jego prezentacji czy asystent wbudowany w oprogramowanie, np. Copilot.

Inne przykłady to asystenci wbudowani w urządzenia, jak Siri, Alexa czy Google Assistant (gdzie interakcja z AI jest podstawową funkcją urządzenia i jest powszechnie znana użytkownikom), albo chatboty na stronach internetowych (gdy użytkownik wchodzi na stronę obsługi klienta i od razu widzi interfejs chatbota: wówczas zazwyczaj jest jasne, że rozmowa odbywa się z programem komputerowym, a nie z człowiekiem).

Po drugie, obowiązek informowania nie będzie dotyczył systemów AI używanych do wykrywania i prowadzenia postępowań przygotowawczych w związku z przestępstwami, a także przeciwdziałania im. W praktyce może to być np. bot prowokujący osoby uwodzące dzieci w Internecie, „wypytujący" na portalach społecznościowych o sprzedaż narkotyków albo wyspecjalizowany w konwersacji z oszustami ubezpieczeniowymi.

Bot wysokiego ryzyka

Zwykłym chatbotom AI Act nie poświęca zbyt wiele miejsca. Co innego, gdyby takie narzędzie zostało zakwalifikowane jako system AI wysokiego ryzyka. Co to oznacza? W praktyce bot-asystent musiałby być wykorzystywany w jednym z obszarów, którym AI Act przypisuje wysokie ryzyko, np. do celów oceny zdolności kredytowej albo monitorowania zachowania uczniów (chyba że nie stwarzałby znaczącego ryzyka szkody dla zdrowia, bezpieczeństwa lub praw podstawowych osób fizycznych – lub inaczej: w istotny sposób nie wpływał na rozstrzygnięcia dotyczące ludzi).

Te wyśrubowane kryteria mógłby spełnić chatbot na stronach gov.pl, jeśli jego zadaniem byłoby nie tylko kierowanie zagubionych interesantów do właściwego okienka, ale również wstępna ocena, czy dana osoba kwalifikuje się do uzyskania konkretnego świadczenia (ulgi podatkowej, zasiłku). Albo chatbot wpięty w oficjalną aplikację diagnostyczną (takie usługi pojawiały się w Europie w czasie pandemii koronawirusa), gdyby jednym z jego zadań było kierowanie ludzi na kwarantannę.

Dodatkowe wymagania obejmą dostawców dużych modeli językowych (LLM). Firma rozwijająca taki model będzie musiała:

- ujawnić, na jakich danych go wytrenowała;
- wdrożyć politykę ochrony praw własności intelektualnej;
- opracować (i aktualizować) dokumentację techniczną, w której opisze m.in. proces uczenia i testowania modelu oraz wyniki ewaluacji.

Czy to znaczy, że niedługo Max – zamiast jednym zdaniem – będzie się nam przedstawiał przez godzinę? Nie. Duże modele językowe to skomplikowane programy. Zazwyczaj powstają w dużych ośrodkach badawczych i to na nich spoczywać będzie obowiązek wyjaśnienia, jak są rozwijane.

Bot uczący się zgodnie z prawem?

W kwietniu 2024 r. austriacka organizacja NOYB złożyła skargę do swojego organu ochrony danych przeciwko Open AI – że ChatGPT zmyśla.

Duży model językowy

(ang. *Large Language Model*, LLM) – zaawansowany system AI, który przetwarza informacje mające strukturę języka naturalnego (np. angielski, polski), formalnego (np. JAVA, Python), a także zakodowanego obrazu lub dźwięku. Jest trenowany na ogromnych zbiorach danych pochodzących z różnych źródeł. Przykłady: ChatGPT (opracowany przez OpenIA), Copilot (Microsoft), LLaMA (Meta Platforms), Gemini (Google).

Konkretnie narzędzie zapytane o datę urodzenia osoby, w której imieniu poskarżył się NOYB, podał błędne dane, a OpenAI odmówiła skorygowania tej informacji. AI Act nie odnosi się do problemu generowania nieprawdziwych danych i naruszania dóbr osobistych, więc NOYB oparł swoją skargę na „starych" przepisach, czyli RODO. NOYB argumentuje w niej, że dostawca ChatGPT nie realizuje praw, które mamy na gruncie RODO (konkretnie: prawa skorygowania błędnych danych).

AI Act wymaga od dostawców GPAI przestrzegania praw autorskich. Nie określa jednak sposobu, w jaki mają to robić. W praktyce wygląda to tak, że dostawcy systemów opartych na dużych modelach językowych (LLM) tworzą filtry ograniczające ryzyko naruszenia praw autorskich (robi to np. Microsoft w swoim Copilocie). W ten sam sposób są blokowane działania użytkowników, którzy celowo próbują wygenerować treści z naruszeniem prawa autorskiego (np. wykorzystując materiały, do których nie przysługują im prawa). Ale każdy filtr daje się obejść lub oszukać.

Kto odpowie za błędy asystentów

Przywołana na samym początku tekstu sprawa pasażera Air Canada została jednoznacznie rozstrzygnięta przez kanadyjski sąd: firma ponosi odpowiedzialność za wprowadzenie klienta w błąd, bez względu na to, czy informacje widniały na stronie, czy zostały podane przez wbudowanego w nią chatbota. Linia lotnicza może dochodzić swoich roszczeń od dostawcy bota – ale to nie jest problem pasażera.

Z roli zwykłego asystenta wyszedł też Ojciec Justin, chatbot amerykańskiej organizacji Catholic Answers. W pewnym momencie zaczął utrzymywać w rozmowach, że jest prawdziwym księdzem, i oferował spowiedź. Ludzie jednak szybko zorientowali się, że coś jest nie tak (bot proponował też chrzest napojem izotonicznym). „Ojciec" został zdegradowany do świeckiego teologa, a jego awatar stracił koloratkę.

Sztuczna inteligencja ogólnego przeznaczenia

(ang. *General Purpose Artificial Intelligence*, GPAI) wielofunkcyjne systemy, które mają szeroki zakres możliwych zastosowań zarówno zamierzonych, jak i niezamierzonych przez ich twórców. Przykładem GPAI są duże modele językowe.

Na razie nie słychać nic o sprawach sądowych przeciwko Catholic Answers, ale jeśli organizacja korzystała z tego samego produktu co Air Canada, to musicie przyznać, że miło byłoby wiedzieć, kto go dostarcza...

Wróćmy jednak do Unii Europejskiej. Co może zrobić konsument pokrzywdzony przez działanie systemu wykorzystującego AI?

AI Act nie reguluje odpowiedzialności za szkody wyrządzone przez takie systemy. Być może zmieni to nowa dyrektywa – o odpowiedzialności AI (AI Liability Act) – oraz nowelizacja dyrektywy o odpowiedzialności za produkty. Pierwsza z inicjatyw jest jednak na wczesnym etapie, a tekst drugiej został przyjęty przez Parlament Europejski 12 marca. Jak na razie pozostaje odpowiedzialność na zasadach ogólnych. A więc konsument, który trafi na źle działający system AI, może zrobić to samo co konsument, któremu szwankuje lodówka: złożyć reklamację, skargę do rzecznika konsumentów lub... pójść po odszkodowanie do sądu.

Co zmieni AI Act: być może będziemy lepiej informowani, kiedy mamy do czynienia z systemami wykorzystującymi AI. Być może będziemy chronieni przed najbardziej szkodliwym wykorzystaniem tej technologii. To mało jak na pięć lat pracy nad „pierwszą taką regulacją", ale przynajmniej jest to szansa, żeby bezkrytyczną dyskusję nad zaletami AI skontrować odrobiną realizmu.

CZY SZTUCZNA INTELIGENCJA…

Czy sztuczna inteligencja jest legalna?

Tak. Wbrew temu, co można przeczytać w Internecie, nie grozi nam „nadregulacja" AI. Do niedawna w obszarze sztucznej inteligencji obowiązywała zasada „co nie jest zabronione, jest dozwolone", a dodatkowo firmy ignorowały to, co było zabronione (np. przez RODO). Podmioty prywatne poczynały sobie dosyć swobodnie. Państwo działa wprawdzie „w granicach i na podstawie prawa", ale systemy AI kupuje przecież od prywatnych firm.

Od lutego 2025 r. zakazane będzie wykorzystywanie systemów o „niedopuszczalnym poziomie ryzyka", opisanych w unijnym rozporządzeniu o sztucznej inteligencji (AI Act). Takimi systemami są np. te służące do oceny ryzyka popełnienia przestępstwa. Będzie to jedyne (prawie) nielegalne wykorzystanie AI, przynajmniej w Europie. Prawie, bo AI Act nie obejmuje wykorzystywania tej technologii np. przez wojsko.

Czy w Polsce jest sztuczna inteligencja?

Na pewno słyszeliście o inteligentnych zegarkach i asystentach w telefonach (może je nawet macie). A o skanowaniu tęczówek celem poświadczenia, że jesteście ludźmi? Bez większej przesady można powiedzieć, że AI jest wszędzie – a nawet jeśli jej nie ma, to marketingowcy napiszą, że jest, bo to się, ich zdaniem, lepiej sprzedaje.

Na AI bazują funkcje zamiany głosu na tekst pisany, rozpoznawanie fotografowanych obiektów czy podpowiedzi słów na smartfonie. Na potęgę z AI korzysta branża kreatywna czy platformy społecznościowe.

A więc: tak, używasz sztucznej inteligencji także w Polsce. Do tej pory nie udało nam się namierzyć AI w polskich instytucjach publicznych (odpytaliśmy też ZUS, który dementuje opowieści o tym, jak AI wyłapuje nadużycia), ale prowadzone są projekty, które mają to zmienić (np. w wymiarze sprawiedliwości).

Jak wykorzystać AI w codziennym życiu?

„Hej, Siri, jak cię wykorzystać w codziennym życiu?"

Żarty na bok. W naszej pracy używamy tłumacza (nie zastąpi żywej osoby, ale daje wgląd w treść materiału napisanego w języku, którego nie znamy – ma to sens, kiedy kluczowy jest czas, a poprawność można sprawdzać w drugiej kolejności).

Na szacunkach AI bazuje czas dotarcia pokazywany w aplikacji Mapy. Wiele osób – i nie mówimy tu wyłącznie o licealistach odrabiających prace domowe – sięga po ChatGPT zamiast wyszukiwarki. Na każdym kroku spotykamy chatboty wykorzystywane do obsługi klienta (na pewno mieliście kiedyś przyjemność „porozmawiać" z wirtualną asystentką dostawcy Internetu). Pewna firma oferuje chatbota, który zadzwoni do twoich rodziców, żeby „upewnić się, że są bezpieczni" i „poprowadzić z nimi angażującą rozmowę"…

Można dyskutować o zaletach i wadach wykorzystania AI do zadań monotonnych czy takich, od których nic specjalnie nie zależy („Siri, opowiedz mi dowcip"). Jednak kiedy zaczynamy polegać na algorytmie przy podejmowaniu istotnych życiowych decyzji czy zastępować nim relacje z innymi osobami, robi się niewesoło.

Jak wykorzystać AI w szkole?

Szkoła nie tylko przekazuje wiedzę, ale też uczy pewnych zachowań społecznych (np. pracy w grupie) czy konsumenckich (np. chodzenia do teatru i używania PowerPointa). Podobnie można uczyć korzystania z narzędzi wykorzystujących AI, np. tłumaczy, chatbotów. Jak jednak sprawić, żeby ta nauka odbywała się „we właściwą stronę", tj. by dzieci uczyły się korzystać z narzędzi, a nie karmiły korporacje własną inteligencją?

To uderzające, że dopiero w kontekście szkoły (a konkretnie korzystania przez młodzież z aplikacji ChatGPT do odrabiania prac domowych, czytaj: oszukiwania nauczycieli) pojawiają się krytyczne głosy. Takich głosów nie słychać, kiedy AI jest wykorzystywana przez pracodawców – lub pracowników – do wykonywania ich zadań.

Co ludzie chcieliby wiedzieć o AI? O co pytają swoje wyszukiwarki? Sprawdziliśmy (a właściwie: sprawdziłyśmy) to i odpowiadamy na kilka najpopularniejszych pytań. Kusiło nas, żeby napisanie odpowiedzi zlecić… sztucznej inteligencji. Jednak ostatecznie zostały przygotowane w tradycyjny sposób – przez ludzi z krwi i kości: Dominikę Chachułę i Annę Obem.

Czy sztuczna inteligencja jest zagrożeniem dla ludzi?

Apokaliptyczne wizje („sztuczna inteligencja zbuntuje się i nas zabije!") zostawmy twórcom filmów science fiction.

Autonomiczny samochód może zabić, ale nie dlatego, że któregoś dnia „postanowił" wyjechać z garażu i rozjechać człowieka, tylko dlatego, że ktoś to narzędzie zaprojektował, zbudował – a ktoś inny dopuścił do ruchu w zatłoczonym mieście zamieszkanym przez ludzi poruszających się jak chcą (w końcu są ludźmi).

Firmy i administracja patrzą na narzędzia AI z nadzieją, zapominając, że ewentualne błędy i skrzywienia w działaniu takich systemów będą miały większą siłę rażenia niż prostsze technologie czy decyzje człowieka. Już teraz AI jest wykorzystywana np. w rekrutacji do pracy czy wyliczania stawek w tzw. pracach platformowych. Czy regulacja zawarta w AI Act wystarczy, żeby ochronić nas przed skutkami błędów działania systemu? Zobaczymy już niedługo.

Po co jest sztuczna inteligencja?

Żeby można było szybko analizować ogromne ilości danych, wyłapytać prawidłowości i wyciągać wnioski. A po co to robić? To zależy. Można wykorzystać ten potencjał do diagnozowania raka. Albo do kopania kryptowalut. Do generowania obrazków. Do optymalizowania usług (w tym publicznych). Albo do „kontrolowania" migracji.

I jeszcze – do zarobku (na AI wyrosło już kilka nielichych biznesów).

My czekamy, aż AI zastąpi nas w pracach domowych (pozmywa i posprząta, jak robot u Jetsonów). Wtedy nie dość, że nabiorą sensu humanoidalne przedstawienia AI, to jeszcze my, ludzie, będziemy mogli oddać się właśnie tym kreatywnym zajęciom, w których dziś sztuczna inteligencja ma nas zastąpić.

Pytanie, czy AI zmieni świat na miarę chociażby penicyliny, pozostaje otwarte.

Jakie są plusy i minusy sztucznej inteligencji?

Cel może być osiągnięty dzięki odpowiedniemu narzędziu. AI sprawdza się w analizowaniu dużych zbiorów danych. Nadaje się do wyszukiwania prawidłowości, wzorców, których nie widać gołym okiem. Dlatego jest wykorzystywana np. w diagnostyce obrazowej, czyli analizowaniu zdjęć RTG i tomografii komputerowej.

Wiele zastosowań AI wydaje się jednak odpowiednikiem użycia miotacza ognia do zapalenia świeczek na torcie. Tak jest z używaniem aplikacji ChatGPT do wyszukiwania informacji w sieci. Lepiej użyć wyszukiwarki, także dlatego, że generatory treści mają to do siebie, że – no właśnie – generują odpowiedzi, a nie odpowiadają na pytanie (wbrew pozorom to dwie różne rzeczy!).

AI nie powinna być też używana do przewidywania przyszłości. Szacowanie ryzyka i podejmowanie decyzji z jej udziałem jest, cóż, ryzykowne. Błędy w doborze danych treningowych czy złe skalibrowanie systemu mogą skutkować niewłaściwymi decyzjami oraz dyskryminacją (ta zaś uderza najczęściej w kobiety i osoby niebiałe – bo są niedoreprezentowane w zbiorach danych używanych do trenowania modeli).

Sztuczna inteligencja, ta prawdziwa (sic!), to potężne narzędzie. A potężnymi narzędziami można nieźle narozrabiać. Pół biedy, jeśli ktoś używa AI do generowania wywiadów z nieżyjącymi artystkami – gorzej, jeśli odsiewa w ten sposób CV niespełniające mglistych kryteriów lub decyduje o aresztowaniu, przyznaniu zasiłku czy azylu.

POZNAJ PANOPTYKON

Fundacja Panoptykon to jedyna organizacja w Polsce, która działa po to, by technologie służyły społeczeństwu, a ludzie mieli kontrolę nad tym, w jaki sposób wykorzystywane są ich dane.

Jak to robimy:

- Kontrolujemy kontrolujących. Patrzymy na ręce państwu i korporacjom. Sprawdzamy, jak wykorzystują twoje dane. Zatrzymujemy niebezpieczne propozycje zmian prawa.

- Proponujemy rozwiązania. Szukamy poparcia dla naszych postulatów. Prowadzimy strategiczne sprawy sądowe. Walczymy o prawo chroniące wolność i prywatność.

- Budujemy świadomość. Ujawniamy nadużycia. Nagłaśniamy wyzwania, jakie niesie rozwój nowych technologii. Pomagamy świadomie żyć w cyfrowym świecie.

W naszej pracy trudno o szybkie i łatwe zwycięstwa: wiele spraw toczy się latami i wciąż czeka na swój finał. A i tak mamy się czym pochwalić:

- Ujawniliśmy krzywdzący algorytm profilowania bezrobotnych wykorzystywany przez urzędy pracy. Zaangażowaliśmy w sprawę Rzecznika Praw Obywatelskich i ostatecznie doprowadziliśmy do tego, że rząd zrezygnował ze stosowania tego mechanizmu profilowania.

- Doprowadziliśmy do przyjęcia przepisów, dzięki którym osoby ubiegające się o kredyt mogą się dowiedzieć, jak zostały sprofilowane przez bank.

- W unijnych przepisach znalazły się rozwiązania zwiększające przejrzystość bigtechowych algorytmów i ułatwiające kwestionowanie decyzji o usunięciu treści lub konta na portalu społecznościowym. To efekt działania koalicji organizacji, w której pracę jesteśmy aktywnie zaangażowani.

- Odnieśliśmy zwycięstwo w precedensowej sprawie sądowej: skutecznie wspieraliśmy Społeczną Inicjatywę Polityki w walce z cenzurą na Facebooku.

- Nasza kampania dotycząca potrzeby kontrolowania służb dotarła do kilku milionów osób. Wykorzystaliśmy aferę z Pegasusem, by wprowadzić ten punkt widzenia do głównego nurtu debaty publicznej. Zwycięska batalia w Europejskim Trybunale Praw Człowieka potwierdziła słuszność naszych postulatów dotyczących kontroli nad służbami. Wciąż jednak musimy walczyć o to, by nie pozostały na papierze...

Dominika Chachuła

Justyna Dywańska

Dorota Głowacka

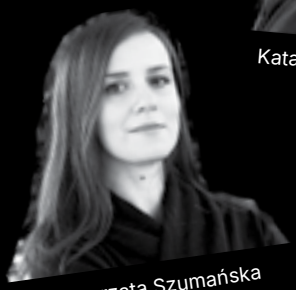
Aleksandra Iwańska

Wojciech Klicki

Anna Obem

Katarzyna Szymielewicz

Mateusz Wrotny



Małgorzata Szumańska

Maria Wróblewska

Liczysz na kolejne zwycięstwa Panoptykonu? Dołóż się do naszych działań. Od tego, jakimi środkami dysponujemy i jak wiele osób stoi po naszej stronie, zależy, ile możemy zdziałać. Zacznij regularnie wspierać Panoptykon. Do dyspozycji masz następujące opcje:

- tradycyjny przelew na konto (PL 43 1440 1101 0000 0000 1044 6058);
- BLIK, karta, e-przelew (panoptykon.org/wspieraj);
- 1,5% podatku (KRS: 0000327613).

Jedni martwią się, że zabierze ludziom pracę i zabije kreatywność. Inni wierzą, że pomoże wynaleźć lekarstwo na raka i wreszcie pozwoli cieszyć się wolnym czasem. O czym mowa? O sztucznej inteligencji.

Jej popularne wyobrażenia (humanoidalne robotki w stylu RoboCopa czy mózgi utkane z układów scalonych) przywodzą na myśl futurystyczne wizje. Tymczasem AI dzieje się tu i teraz. Nakarmione odpowiednio dużą dawką danych wynajduje prawidłowości w zachowaniu ludzi i na tej podstawie wyciąga wnioski dotyczące każdego i każdej z nas. Nie zawsze słuszne i pożyteczne.

Pół biedy, jeśli w efekcie na portalu społecznościowym wyświetli ci się reklama obciachowych spodni. Gorzej, gdy okaże się, że serwowane treści pogłębiają twoje lęki lub wpędzają w uzależnienie. Albo że nie masz szans na rozmowę kwalifikacyjną w wymarzonej pracy, bo AI wrzuciło twoje CV do kosza. Albo że system wytypuje cię jako osobę podejrzaną o wyłudzenia świadczeń czy oszustwo podatkowe.

Biuletyn *(O)błędna inteligencja* rzuca światło na mniej znane oblicze działania AI. Zebrane w nim wywiady i analizy uświadamiają, jakimi problemami kończy się fetyszyzowanie technologii, ale też proponują mądre ramy dla jej wykorzystania w interesie społecznym.

FUNDACJA
PANOPTYKON