

Dyskryminacja w świecie rządzonej przez dane

28 lipca 2018

Technologie oparte na danych nie są neutralne. Decyzje o zbieraniu, analizowaniu i przetwarzaniu konkretnych kategorii informacji są dyktowane czynnikami społecznymi, ekonomicznymi i politycznymi. Takie operacje z wykorzystaniem danych osobowych mogą nie tylko naruszać prywatność, ale też prowadzić do dyskryminacji i pogłębienia wykluczenia już marginalizowanych społeczności. Dyskryminacja, do której prowadzi przetwarzanie danych i wykorzystanie algorytmów, to ogromne wyzwanie dla obrońców praw człowieka, a odpowiedź na nie nie jest oczywista. Raport pt. „Between Anti-discrimination and Data” („Między przeciwdziałaniem dyskryminacji a danymi”) próbuje rzucić na problem nieco światła z perspektywy europejskich organizacji społecznych.

Opresyjne dane

Dyskusja o potencjalnym wpływie systemów informatycznych na prawa człowieka wiąże się z historią nadzoru nad grupami historycznie marginalizowanymi. Przez stulecia różne instytucje tworzyły rejestry identyfikujące osoby i grupy „niepożądane”, „nienormalne” i „niebezpieczne”. Zazwyczaj były to grupy, które nazwalibyśmy dziś marginalizowanymi społecznie (np. bezdomni, prostytutki, mniejszości religijne). Zarówno totalitarne, jak i demokratyczne rządy wykorzystywały spisy ludności do namierzania mniejszości etnicznych, rasowych, językowych czy rdzennej ludności, co prowadziło do fatalnych skutków: zbrodni przeciwko ludzkości, ludobójstwa, wysiedleń. Wśród najbardziej ekstremalnych przypadków, do których doprowadziło niewłaściwe wykorzystanie informacji, historia wymienia holokaust, rewolucję kulturalną w Chinach, apartheid w południowej Afryce czy ludobójstwo w Rwandzie.

Ale są też mniej odległe przykłady. W 2013 r. szwedzka prasa ujawniła, że policja z regionu Skåne prowadziła elektroniczny rejestr osób o pochodzeniu romskim. Baza obejmowała informacje o 4 tys. osób, z czego tysiąc to dzieci. Oficjalnie rejestr miał pomagać w zwalczaniu przestępczości i prowadzeniu polityki wobec migrantów. Do rejestru dostęp miało przynajmniej 70 funkcjonariuszy. Afera zakończyła się likwidacją rejestru, a dzięki interwencji jednej z organizacji społecznych niektóre osoby dotknięte negatywnymi konsekwencjami otrzymały odszkodowanie.

Algorytmiczna dyskryminacja

Powyższe przykłady dotyczą mało zaawansowanych technologicznie narzędzi opresji. Gwałtowna transformacja w przechowywaniu, przetwarzaniu i przekazywaniu danych prowadzi do poważniejszych zagrożeń: dyskryminacji zaawansowanej technologicznie (hi-tech). Instytucje i firmy potrzebują coraz bardziej wyrafinowanych metod do analizowania gigantycznej ilości zgromadzonych danych. Z pomocą przychodzą tutaj algorytmy i modele statystyczne. Pozwalają one na kategoryzację, poszukiwanie korelacji, regularności i schematów, a także zautomatyzowanego podejmowania decyzji.

Wykorzystywane techniki mają różny charakter, cele i poziom zaawansowania. Banki na przykład używają algorytmów, żeby zmierzyć zdolność kredytową klientów. Pomoc społeczna z ich pomocą próbuje walczyć z wyłudzaniem zasiłków. Dominująca narracja podkreśla przydatność narzędzi w unowocześnianiu procesu podejmowania decyzji – jeśli chodzi o efektywność i czasochłonność. Jednakże postęp techno-analityczny ma też swoją ciemną stronę. Cały czas toczy się dyskusja o tym, czy algorytmy mogą prowadzić do dyskryminacji i wykluczenia społecznego.

W tym kontekście pojawiają się dwa problemy. Pierwszy ma związek ze źle dobranym zestawem danych, danymi

niekompletnymi, nieaktualnymi lub takimi, przez które niektóre grupy społeczne są mniej widoczne. Jak podkreśla K. Crawford, dane nie są zbierane tak samo od wszystkich. Zdarza się, że całe grupy lub pojedyncze osoby wymykają się procesowi zbierania informacji z powodu sytuacji finansowej, miejsca zamieszkania, umiejętności korzystania z komputerów, stylu życia. Dane mogą też powielać historyczne uprzedzenia. Jeden z wczesnych przykładów dotyczy brytyjskiej St. George Medical School, w której zautomatyzowano proces rekrutacji kandydatów zainteresowanych edukacją w tej konkretnej placówce. System rekrutacji studentów powstał w oparciu o powstały dawniej model rekrutacji kandydatów do pracy i zawierał mechanizm odpowiedzialny za przyznawanie gorszej punktacji wnioskowi złożonemu przez kobiety i osoby o nazwiskach innych niż europejskie.

Niepewna jakość danych wejściowych to nie jedyna obawa. Drugim poważnym problemem jest konstrukcja algorytmów. Tworzenie modeli analitycznych należy do ludzi. Proces ten zazwyczaj odzwierciedla poglądy osób programujących na to, jak powinien wyglądać świat. Człowiek decyduje o rodzajach błędów i wskaźnikach dla modeli algorytmicznych. Innymi słowy: ktoś musi zdecydować, czy mierzyć „wiarygodność” algorytmu względem przypadków, wobec których została podjęta nieprawidłowa decyzja pozytywna (false positives) lub negatywna (false negatives). Ktoś musi też zdecydować, jaki jest akceptowalny poziom nieprawidłowych decyzji pozytywnych i negatywnych.

Jeden z najdobitniejszych przykładów dotyczy szeroko dyskutowanej sprawy amerykańskiego algorytmu do orzekania wyroków, o której napisał portal „Pro Publica”. System ten ocenia ryzyko popełnienia przez podejrzanego podobnego czynu w przyszłości i na tej podstawie podejmowana jest decyzja o tymczasowym aresztowaniu. Dziennikarze „Pro Publici” wykryli, że system może traktować osoby czarne gorzej niż białe, podejrzone o taki sam czyn.

Pudełko z błędami

Do nierównego traktowania w związku z przetwarzaniem danych może dochodzić w różnych sytuacjach. Bardzo często będzie to nawet trudno wykryć. To czarna skrzynka (black box) – nietransparentny mechanizm, który wykonuje niezrozumiałe operacje na danych. Brak przejrzystości oznacza, że osoby dotknięte taką decyzją (dotyczącą tego, że np. nie dostaną kredytu, nie zostaną przyjęte na uczelnię, nie dostaną pracy) mają znikomą szansę dowiedzenia się, co się wydarzyło, i zakwestionowania wyniku.

Dyskryminujące i niebezpieczne czarne pudełka bardzo rzadko powstają w wyniku celowych działań podjętych przez człowieka. Programiści czasem się mylą. Doskonałego przykładu takiej sytuacji dostarcza system wykorzystywany w amerykańskim stanie Michigan. Agencja zatrudnienia wprowadziła zautomatyzowany system do wykrywania nadużyć. System działał w latach 2013–2015 – i był pełen błędów. Ze względu na złe zaprojektowanie wygenerował ponad 20 tys. wskazań potencjalnych podejrzanych, wstrzymał wypłatę zasiłków dla bezrobotnych i zarekomendował kary na łączną kwotę 100 tys. dolarów. Przeciwko administracji stanowej złożono w 2015 r. liczne pozwy.

Wyzwanie dla społeczeństwa obywatelskiego?

Opresyjne zbieranie danych i dyskryminujące algorytmy mogą prowadzić do pogwałcenia praw człowieka na masową skalę. Żeby temu zapobiegać, organizacje broniące praw człowieka powinny poszukać niekonwencjonalnych rozwiązań. Wykrywanie i neutralizacja szkodliwych zdarzeń z wykorzystaniem danych wymagają od sektora społecznego zarówno ekspertyzy technicznej, jak i wrażliwości na sprawiedliwość społeczną. Te dwie cechy jednak nie zawsze idą w parze. Bardzo często

organizacje operują w silosach – osobno traktując prawa cyfrowe, przeciwdziałanie dyskryminacji i obronę praw mniejszości. W dobie dyskryminacji powodowanej przez algorytmy te siły trzeba łączyć.

W naszym raporcie rekomendujemy kilka sposobów walki z problemem dyskryminacji powodowanej przez algorytmy. Dostrzegamy potencjał pracy z dziennikarzami śledczymi, sprytne wykorzystanie obowiązującego od niedawna RODO, budowanie koalicji między różnymi sektorami. Nieodzowne jest też tworzenie nowej narracji i proponowanie strategii łączenia prywatności w środowisku cyfrowym i wrażliwości na niesprawiedliwość społeczną.

Dostrzegamy trzy sposoby wsparcia organizacji w zajęciu się problemem zautomatyzowanej dyskryminacji:

- wsparcie aktywistów zajmujących się prawami cyfrowymi i prywatnością w zrozumieniu problemu dyskryminacji i połączeniu go z problemem ochrony danych;
- wsparcie organizacji zajmujących się problemem dyskryminacji i sprawiedliwością społeczną w dostrzeżeniu związku między nierównościami społecznymi i technologiami opartymi na przetwarzaniu danych;
- wsparcie elastycznego podejścia w przeciwdziałaniu zautomatyzowanej dyskryminacji (połączenie narzędzi zarówno z zakresu praw cyfrowych, jak i prawa antydyskryminacyjnego).

Tłumaczenie: Anna Obem

Źródło: Panoptikon.org

BIBLIOGRAFIA

1. <http://eprints.lse.ac.uk/88053/>

2.

http://www.repository.fedesarrollo.org.co/bitstream/handle/11445/1093/Co_So_Junio_2005_Seltzer.pdf

3 .

<https://www.opendemocracy.net/can-europe-make-it/mattias-gardell/swedens-dirty-little-secret>

4 .

<http://europepmc.org/backend/ptpmcrender.fcgi?accid=PMC2545288&blobtype=pdf>

5 .

<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

6 .

<http://blogs.lse.ac.uk/mediapolicyproject/2016/02/05/bittersweet-mysteries-of-machine-learning-a-provocation/>

7 .

<https://www.theguardian.com/us-news/2016/dec/18/michigan-unemployment-agency-fraud-accusations>