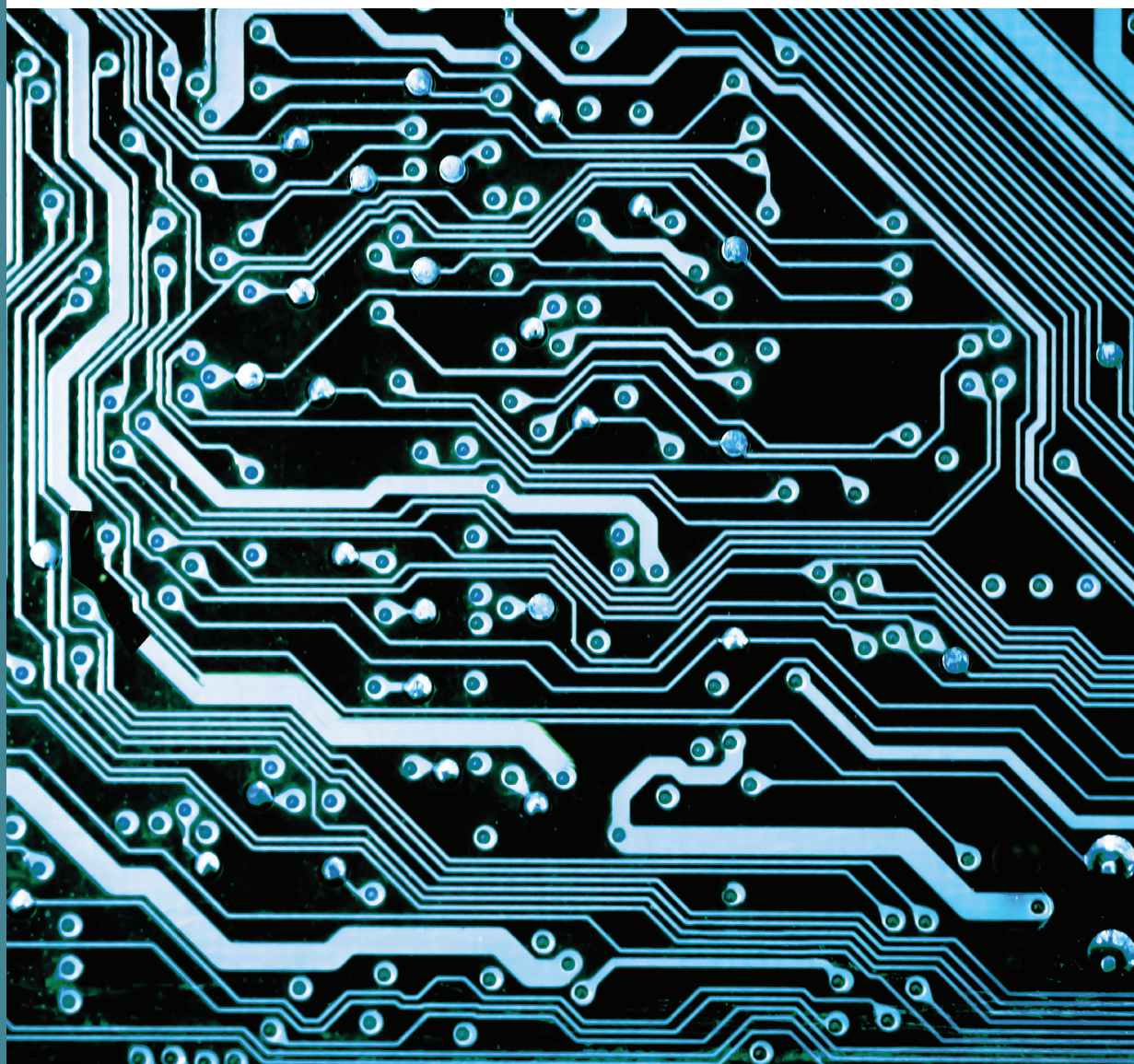




Natalia Mileszyk, Bartosz Paszcza, dr Alek Tarkowski

AlgoPolska

Zautomatyzowane podejmowanie decyzji
w służbie społeczeństwu



AlgoPolska

Zautomatyzowane podejmowanie decyzji
w służbie społeczeństwu

Natalia Mileszyk

Bartosz Paszcza

dr Alek Tarkowski

Sfinansowano przez Narodowy Instytut Wolności -
Centrum Rozwoju Społeczeństwa Obywatelskiego
ze środków Programu Rozwoju Organizacji Obywatelskich na lata 2018-2030



Narodowy Instytut Wolności
Centrum Rozwoju Społeczeństwa Obywatelskiego



Program Rozwoju
Organizacji
Obywatelskich
na lata 2018-2030

PROO

Redakcja: Paweł Musiałek
Projekt graficzny: Rafał Gawlikowski
Korekta: Marta Wojaś
Skład: Maciej Kielbas
Grafika na s. 18: Magdalena Karpińska

ISBN 978-83-956101-2-7

Fundacja Centrum Cyfrowe
Klub Jagielloński

Kraków, Warszawa 2019

Ten utwór jest udostępniony na licencji Creative Commons Uznanie Autorstwa 4.0 Międzynarodowe. Zachęcamy do jego przedruku i wykorzystania. Prosimy jednak o zachowanie informacji o finansowaniu raportu oraz podanie linku do naszej strony.

Autorzy dziękują uczestnikom seminarium dotyczącego raportu: Joannie Mazur, Pawłowi Musiałkowi, Michałowi Palińskiemu, Annie Podgórskiej-Buompane, Katarzynie Szymielewicz, Zuzannie Warso, Krzysztofowi Wojdyle, Janowi Zygmuntowskiemu.

Natalia Mileszyk

Ekspertka polityk publicznych w Fundacji Centrum Cyfrowe. Prawniczka specjalizująca się w prawach cyfrowych, reformie praw autorskich i polityce otwartości. Członkini Stowarzyszenia Communia, gdzie koordynuje zespół adwokatów, ekspertka prawna w Creative Commons Polska. Przez ostatnie trzy lata była aktywnie zaangażowana w reformę praw autorskich na poziomie europejskim. Absolwentka Uniwersytetu Warszawskiego i Uniwersytetu Środkowoeuropejskiego w Budapeszcie (LL.M.).

Bartosz Paszcza

Ekspert do spraw nowych technologii Centrum Analiz Klubu Jagiellońskiego, redaktor portalu klubjagiellonski.pl, magister studiów nad internetem Uniwersytetu w Southampton. Jest współzałożycielem Fundacji Polonium, organizacji łączącej polskich naukowców za granicą. Członek zarządu Klubu Jagiellońskiego.

dr Alek Tarkowski

Prezes i współzałożyciel Fundacji Centrum Cyfrowe. Od 2004 roku jest aktywny w Polsce i na świecie w organizacjach i ruchach społecznych budujących otwarty internet. Jest współzałożycielem Creative Commons Polska, Stowarzyszenia Communia i Polskiej Koalicji na rzecz Otwartej Edukacji (KOED). Członek zarządu organizacji Creative Commons i komitetu sterującego Internet Governance Forum Polska.

SPIS TREŚCI

8 / Streszczenie

10 / Wprowadzenie

- 10 / Historia XXI wieku – od internetu do sztucznej inteligencji**
- 11 / Natura rewolucji AI – pomiędzy R2D2 a rzeczywistością**
- 12 / Automatyczne podejmowanie decyzji jest ważniejsze od AI**
- 13 / Technologiczny żabi skok dzięki debacie publicznej**
- 14 / Popyt warunkiem zrównoważonego rozwoju ADM**
- 15 / Zaleta zapóźnienia: mądry Polak po czyjejsz szkodzie**
- 16 / Polska droga do technologii w służbie społeczeństwa**

17 / Systemy ADM – przykłady z życia

- 17 / Charakterystyka systemów zautomatyzowanego podejmowania decyzji**
- 19 / Prewencja przestępstw**
- 21 / Rynek pracy**
- 22 / Opieka społeczna**
- 23 / Opieka medyczna**
- 25 / Edukacja**
- 26 / Dostęp do informacji i kultury**
- 28 / Finanse**
- 29 / Sądownictwo**

31 / Wyzwania i szanse związane z ADM

31 / Szanse

- 31 / Większa efektywność**
- 33 / Podejmowanie decyzji w oparciu o dane i wiedzę**
- 34 / Obserwacja i eliminacja części ludzkich błędów i stronniczości**

34 / Wyzwania

- 34 / Współpraca człowiek–algorytm**
- 36 / Sprawiedliwość**

- 39 / Przejrzystość i wytłumaczalność**
- 40 / Globalizacja pomijająca lokalne wartości**
- 41 / Prywatność**
- 42 / Odpowiedzialność**

45 / Rekomendacje

- 46 / Obszary i propozycje rozwiązań**
 - 46 / Realna debata i współpraca – model poczwórnej helisy**
 - 47 / Model regulacji**
 - 49 / Wyjaśnienia, audyt i możliwość odwołania od decyzji**
 - 51 / Zarządzanie danymi**
 - 52 / Edukacja**

54 / Kluczowa literatura

Streszczenie

- Proponujemy przeniesienie punktu ciężkości debaty z dyskusji o sztucznej inteligencji na dyskusję o zautomatyzowanych systemach podejmowania decyzji (*Automated Decision Making*, ADM). Systemy ADM łączą infrastrukturę techniczną i procedury społeczne, by delegować na komputerowe modele część decyzyjności będącej dotychczas w rękach ludzi. Często, ale nie zawsze, modele te wykorzystują sztuczną inteligencję – technologię uczenia się maszynowego.
- Mówienie o systemach ADM – zamiast o sztucznej inteligencji – pozwala lepiej rozumieć wpływ tych nowych technologii na jednostkę i społeczeństwo. Automatyzacja podejmowania decyzji jest też bowiem wprowadzana z wykorzystaniem dużo prostszych technologii.
- Gospodarczo-technologiczny skok Polski nie dokona się bez wytworzenia popytu na wdrażanie najnowszych technologii w kraju. Obserwujemy skupienie debaty oraz polityki rozwoju sztucznej inteligencji w Polsce wokół prac badawczo-rozwojowych oraz wykorzystywania jej do tworzenia eksportowych produktów gospodarczych. Prowadzi to do stworzenia strategii podażowej, w której Polska jest krajem dostarczającym opartych o sztuczną inteligencję rozwiązań stosowanych następnie w innych miejscach na świecie. Stworzenie strategii popytowej jest niezbędne, jeśli rewolucja technologiczna ma przynieść korzyść polskiemu społeczeństwu.
- W Polsce wdraża się mniej systemów ADM niż w innych państwach europejskich i rozwiniętych gospodarkach. Możemy wykorzystać rentę zapóźnienia, by zapewnić zrównoważony rozwój systemów ADM w Polsce. Wprowadzając te systemy, możemy bowiem wyciągać wnioski i uczyć się na błędach popełnianych w innych państwach, które dokonały pionierskich – i często wadliwych – wdrożeń. Rozsądne, określone jasnymi zasadami dbałości o interes publiczny, wprowadzanie systemów ADM może przynieść korzyści polskiej gospodarce, społeczeństwu i administracji publicznej.
- W raporcie prezentujemy wybór różnych zastosowań systemów ADM na świecie, w Europie i Polsce w kluczowych dla życia społecznego obszarach: prewencji przestępstw, rynku pracy, opiece społecznej i medycznej, edukacji, dostępie do kultury i informacji, finansach i sądownictwie. Wskazujemy też pozytywne skutki, jakie niesie za sobą wdrażanie systemów automatyzujących podejmowanie decyzji: podniesienie efektywności działania przedsiębiorstw i administracji, podejmowanie decyzji w oparciu o dane i wiedzę, identyfikację i eliminację części ludzkich błędów oraz stronniczości.

- Równocześnie opisujemy wyzwania związane z wdrażaniem systemów ADM: zbyt wielkie zaufanie do algorytmicznych decyzji skutkujące utratą samodzielności przez ludzi, przyczynianie się systemów ADM do społecznych niesprawiedliwości i nierówności, brak przejrzystości algorytmów i związany z tym brak odpowiedzialności, globalizację systemów ADM prowadzącą do utraty wrażliwości na wartości lokalnych społeczności, ochronę prywatności i innych praw podstawowych.
- Jeśli chcemy budować społeczne i gospodarcze zaufanie do systemów ADM, konieczna jest współpraca w modelu poczwórnej helisy, łączącym naukowców, przedsiębiorców, administrację publiczną i stronę społeczną. Jedynie uwzględniając interesy każdej z grup, będziemy w stanie automatyzować procesy społeczne w sposób zrównoważony.
- Potrzebna jest także regulacja systemów ADM. Ze względu na wpływ, który systemy te mają na jednostki i całe społeczeństwo, konieczne jest wspólne wypracowanie założeń regulacyjnych mających zapewnić, że systemy ADM będą działać w łączonym interesie społeczeństwa oraz gospodarki. Regulacje muszą przede wszystkim zapewnić ochronę praw podstawowych i wskazać obszary, w których systemy ADM nie mogą być stosowane – decyzje muszą być dokonywane przez ludzi.
- Systemy ADM będą jedynie tak dobre, jak wykorzystywane przez nie dane. Odpowiednia jakość oraz przede wszystkim dostępność zbiorów danych jest niezbędna, jeśli chcemy unikać skrzywionych i stronniczych wyników działania systemów ADM. Dostępność danych jest także warunkiem kluczowym do porównawczych analiz wyników działania algorytmów.
- Fundamentem budowania systemów ADM musi być przejrzystość i wytłumaczalność ich działania. Ma to szczególne znaczenie w przypadku systemów stosowanych przez administrację publiczną oraz decydujących o kwestiach wrażliwych, m.in. wpływających na prawa podstawowe obywateli. Pozwala bowiem zrozumieć oddziaływanie tych systemów zarówno w jednostkowych przypadkach, jak i na poziomie społeczeństwa.
- Podstawą zrównoważonego rozwoju ADM musi być świadomość społeczna: zrozumienie działania tych systemów i ich skutków. W społeczeństwie kształtowanym coraz bardziej przez algorytmy koniecznością jest podstawowa znajomość zasad działania algorytmów oraz umiejętność ich wykorzystania ze świadomością tworzonych przez nie szans i zagrożeń. Niezbędną specjalistyczną wiedzę muszą też posiadać osoby odpowiedzialne za wdrażanie systemów ADM. Dlatego edukację – często wspomnianą, choć równie często bagatelizowaną – uznajemy za kluczowy element poszukiwania rozwiązań.

Wprowadzenie

Historia XXI wieku – od internetu do sztucznej inteligencji

Początek XXI wieku był czasem dyskusji nad wpływem technologii na dostęp do treści. Kluczowe spory, nie tylko regulacyjne, lecz także związane z modelami biznesowymi, dotyczyły funkcjonowania prawa autorskiego online, kształtu obiegu treści w sieci. Apogeum tego sporu była kwestia porozumienia handlowego ACTA¹, które doprowadziło w Polsce w 2012 roku do masowych protestów społecznych w imię wolności w internecie.

W drugiej dekadzie XXI wieku ważniejsze stały się jednak dyskusje o technologiach podejmujących decyzje, które mają wpływ na jednostki i społeczeństwo. Debaty te dotyczą z jednej strony doboru i filtrowania treści w serwisach społecznościowych, zaś z drugiej – decyzji podejmowanych przez systemy, które gromadzą i przetwarzają dane o społeczeństwie. Momentami symbolicznego przełomu w debacie publicznej były sprawa Edwarda Snowdena oraz późniejsza afera wokół Cambridge Analytica. Rosnące znaczenie mają kwestie dotyczące zbierania, przepływu i wykorzystywanie danych (głównie osobowych) niezbędnych do podejmowania decyzji.

„Internet” nie jest już jedynym hasłem, które w debacie publicznej łączy kluczowe wątki dotyczące technologii, ich regulacji oraz wpływu na społeczeństwo. Od kilku lat kluczowym pojęciem stała się „sztuczna inteligencja”, która – podobnie jak dwie dekady wcześniej „internet” – opisuje nowy system społeczno-technologiczny kształtujący nasz świat. W obu przypadkach najważniejsza w kształtowaniu polityki publicznej nie powinna być jednak sama technologia, lecz wzajemne oddziaływanie technologii i społeczeństwa.

Internet i AI są przykładami tak zwanych technologii ogólnego zastosowania (*General Purpose Technology*), a więc takich, które są wszechobecne i stają się narzędziem czy bazą dla kolejnych innowacji.

Sztuczna inteligencja, tak jak wcześniej internet, stopniowo wywiera wpływ na wszystkie sfery życia jednostek, także tych niekorzystających z niej bezpośrednio. Możemy więc mówić o epoce sztucznej inteligencji, która nastąpiła po epoce internetu.

² Anti-Counterfeiting Trade Agreement (ACTA), pol. Umowa handlowa dotycząca zwalczania obrotu towarami podrabianymi.

Natura rewolucji AI – pomiędzy R2D2 a rzeczywistością

W debacie o każdej nowej, potencjalnie rewolucyjnej technologii rzeczywistość miesza się z wyobrażeniami. W przypadku internetu w fazie jego wczesnego rozwoju często mylono rzeczywistą sieć komputerową z wyobrażeniami o cyberprzestrzeni zakorzenionymi w literaturze i filmach science fiction². Mówiąc o „społecznościach wirtualnych”, zamiast popularnych wówczas chat roomów, często wyobrażaliśmy sobie całe trójwymiarowe światy.

Dzisiaj, używając pojęcia sztuczna inteligencja, mamy na myśli nie tylko realne technologie oparte na uczeniu się maszynowym, ale też technologie wyobrażone – takie jak inteligentne roboty czy wszechpotężna ogólna sztuczna inteligencja, która może zawładnąć światem. To postaci znane z filmów, takie jak R2D2 z *Gwiezdnych wojen* czy Hal 9000 z *Odysei kosmicznej*. Zjawiska, które przykuwają uwagę mediów i ich odbiorców, często nie są tymi najważniejszymi, czyli mającymi realny wpływ na życie jednostek i społeczeństwa. Mówiąc obrazowo, powinniśmy mniej uwagi poświęcać terminatorowi, a więcej inteligentnemu głośnikowi w naszej kuchni.

Spektakularnym tego przykładem jest robot Sophie, humanoidalna maszyna, która otrzymała obywatelstwo Arabii Saudyjskiej, a w Polsce była gościem specjalnym konferencji Impact CEE. Sophie jest celebrytką świata robotów, zaproszono ją jako realny przykład możliwości oferowanych przez technologie sztucznej inteligencji. Jednak eksperci podważają faktyczny poziom inteligencji i autonomii robota. Yann LeCun, szef badań AI w Facebooku, nazwał Sophie „potiomkinowską AI”³.

Skupiająca uwagę mediów Sophie i jej wizyta w Polsce przesłaniają proces rosnącego znaczenia różnych zastosowań sztucznej inteligencji. Zazwyczaj nie są to jednak humanoidalne roboty, tylko algorytmy o różnej zdolności uczenia się, powodujące automatyzację pracy i procesów podejmowania decyzji w najróżniejszych sferach życia. Przykładem zastosowania takich algorytmów jest trend zrobotyzowanej automatyzacji procesów (*Robotic Process Automation, RPA*) w przedsiębiorstwach. Powoduje on automatyzację różnych obszarów pracy biurowej, na przykład księgowości (piszemy o nich więcej w rozdziale trzecim). Innym przykładem są algorytmy stosowane przez serwisy społecznościowe do kształtowania obiegu treści, a wraz z nimi – sfery publicznej w niemal wszystkich państwach.

Wieloznaczność pojęcia sztuczna inteligencja utrudnia debatę regulacyjną. Jest barierą dla zrozumienia roli tej nowej technologii i jej wpływu na społeczeństwo oraz dla konstruowania polityki publicznej odpowiednio kształtującej jej rozwój. To dlatego, że osoby odpowiedzialne za regulację technologii oraz interesariusze tego procesu mają w głowach wyobrażenia sztucznej inteligencji niejednokrotnie daleko odbiegające od prawdziwego charakteru postępującej transformacji.

² *Cyberpower: the culture and politics of cyberspace and the Internet*, <https://www.taylorfrancis.com/books/9780203448632>

³ *Facebook's head of AI really hates Sophia the robot (and with good reason)*, <https://www.theverge.com/2018/1/18/16904742/sophia-the-robot-ai-real-fake-yann-lecun-criticism>

Automatyczne podejmowanie decyzji jest ważniejsze od AI

Zamiast dyskusji o sztucznej inteligencji proponujemy dyskusję o zautomatyzowanych systemach podejmowania decyzji. Sztuczna inteligencja to przede wszystkim – lub tylko – technologia, co utrudnia debatę regulacyjną i strategiczne podejście do jej rozwoju w kontekście wpływu wywieranego na społeczeństwo. Natomiast systemy ADM to socjotechniczne procesy delegacji części decyzyjności na komputery, niejednokrotnie wykorzystujące AI. To one urzeczywistniają wpływ, który technologia ma na jednostkę i społeczeństwo.

Warto podkreślić, że dyskusja o ADM jest również szersza niż sama debata o sztucznej inteligencji.

Systemy zautomatyzowanego podejmowania decyzji rozumiemy zgodnie z definicją przedstawioną przez Fundację Algorithm Watch jako „procedury, w których podejmowanie decyzji jest początkowo delegowane na inną osobę lub instytucję, która następnie wykonuje tę czynność z pomocą automatycznych modeli podejmowania decyzji”⁴.

W debacie o tych systemach używa się także synonimicznie pojęcia algorytmiczne systemy podejmowania decyzji (*Algorithmic Decision Making*). To pojęcia w dużej mierze zamienne. W raporcie stosujemy termin „zautomatyzowane podejmowanie decyzji”, by podkreślić, że skutki (czyli automatyzacja) są ważniejsze od rozwiązań technicznych (czyli algorytmów).

Systemy ADM są więc stosowane jako nazwa całego procesu podejmowania decyzji, którego elementem, obok człowieka, stają się algorytmy zapisane w kodzie komputerowym charakteryzujące się o różnym poziomem zaawansowania i złożoności. Technologie sztucznej inteligencji są tylko jednym z możliwych narzędzi wykorzystywanych do tworzenia automatycznych modeli podejmowania decyzji. W niektórych wypadkach stosowane algorytmy są bardzo proste i nie korzystają ze sztucznej inteligencji, wciąż jednak stanowią element społeczno-technologicznego systemu zautomatyzowanego podejmowania decyzji.

Skupiając się na systemach ADM, chcemy przesunąć ciężar debaty z dyskusji o technologicznym narzędziu, jakim jest sztuczna inteligencja, do analizy społecznych skutków delegacji decyzyjności na algorytmy. Termin ADM pozwala zauważyć, że proces automatyzacji podejmowania decyzji ma skutki społeczne niezależnie od tego, czy wykorzystuje wysoce zaawansowane, ewoluujące sieci neuronowe, czy prosty zespół sztywno zakodowanych reguł. Debata o ADM przesuwaa nacisk z kwestii rozwoju technologii i gospodarki na kwestię społecznego nadzoru nad tym rozwojem.

⁴ Automating Society – Taking Stock of Automated Decision-Making in the EU, <https://algorithmwatch.org/en/automating-society/>

Technologiczny żabi skok dzięki debacie publicznej

Wdrożenia nowych technologii, a do tych należą systemy ADM, dają szansę na nowe otwarcie – wykorzystanie szans i możliwości kolejnej fali zmiany technologicznej. Odpowiednia strategia, wprowadzona we wczesnej fazie kolejnych zmian technologicznych daje szansę na tzw. żabi skok – radykalną zmianę społeczną i innowacje dzięki nowym wdrożeniom. Strategia żabiego skoku zakłada pominięcie pewnej fazy rozwojowej poprzez przyjęcie strategii szybkiej, radykalnej zmiany. Flagowym przykładem takiej strategii jest Estonia, która w latach 90. XX wieku przekształciła się w społeczeństwo cyfrowe mimo wcześniejszych opóźnień w tej dziedzinie⁵. W Polsce taką strategię przyjął sektor bankowy. Wdrożenia najnowszych rozwiązań bankowości elektronicznej były łatwiejsze dzięki zapóźnieniu technologicznemu sektora.

Zapóźnienie może być powodem troski z perspektywy gospodarczej, jest natomiast szansą na zapewnienie zrównoważonego wykorzystania ADM w Polsce. Renta zapóźnienia – niewielka liczba wdrożonych w Polsce systemów ADM – może w tym wypadku być wykorzystana do zapewnienia społecznej kontroli nad wpływem technologii na jednostkę i społeczeństwo. Ucząc się na błędach popełnianych przez innych i wprowadzając mechanizm społecznej regulacji systemów ADM, mamy szansę podnieść społeczne zaufanie do takich rozwiązań, co jest warunkiem koniecznym do postępu.

Jednocześnie Polska ma małe szanse, by stać się liderem rozwoju technologii sztucznej inteligencji – a więc także systemów ADM. Dzieje się tak mimo naszych ambicji i mimo dużego potencjału polskich startupów AI. Polska powinna więc przyjąć inną strategię, wykorzystując rentę zapóźnienia i realnie oceniając nasz potencjał rozwoju technologicznego. Wraz z popularyzacją systemów ADM będą one wdrażane w wielu państwach, które – podobnie jak Polska – nie są potęgami gospodarczymi czy badawczymi w zakresie rozwoju nowych technologii. Naszą ambicją powinna być wiodąca rola w wyznaczaniu standardów zrównoważonego rozwoju systemów ADM.

W tym celu powinniśmy skupić się nie tylko na rozwoju wąskiego sektora AI w ramach branży informatycznej, ale szerszego systemu społeczno-technologicznego skupionego wokół tych nowych technologii i działającego w sposób zrównoważony. Powinniśmy też zapewnić, że innowacje technologiczne i rozwój gospodarczy będą szły w parze z troską o rozwój społeczny, dobrobyt mieszkańców i redukcję potencjalnych szkód społecznych związanych z wdrażaniem nowych technologii.

Zrównoważony system technologii ADM wymaga wzajemnej komunikacji między czterema środowiskami: tworzącymi rozwiązania, wdrażającymi rozwiązania na potrzeby polskiej gospodarki, wykorzystującymi rozwiązania oraz adaptującymi się do zmian, które systemy ADM powodują w społeczeństwie.

Należy przy tym pamiętać, że narracja radykalnej zmiany, związana z rozwojem sztucznej inteligencji, nie jest prawdziwa. Zarówno wyzwania, jak i ich rozwiązania

⁵ Estonia's digital transformation: Mission mystique and the hiding hand, https://www.ucl.ac.uk/bartlett/public-purpose/sites/public-purpose/files/iipp-wp-2018-09_estonias_digital_transformation.pdf

są podobne do tych, z którymi mieliśmy do czynienia na poprzednich etapach cyfryzacji społeczeństwa. Systemy ADM będą rozwijane w systemie społeczno-technologicznym budowanym w Polsce co najmniej od lat 90. XX wieku. Przykładowo: przy tworzeniu modelowych rozwiązań w zakresie ADM należy opierać się na wcześniejszych debatach i decyzjach regulacyjnych dotyczących np. zarządzania danymi, przejrzystości działań administracji publicznej czy odpowiedzialności za rozwój technologii. Podobnie jest w przypadku działań edukacyjnych niezbędnych dla rozwoju systemów ADM (czy szerzej technologii sztucznej inteligencji). Nie wymagają one radykalnie nowego podejścia – powinny być kontynuacją dobrych strategii i praktyk wskazywanych od lat.

Perspektywa skupiona na społecznym wymiarze technologii pozwala przejść od pytania o to, którą technologię i jak chcemy rozwijać, do pytania: po co to robimy? Dzięki temu sam fakt istnienia danej technologii czy potencjał gospodarczy jej wdrożenia przestaje być decydujący. Istotne stają się pytania o to, jaki będzie zysk dla społeczeństwa i jaki koszt społeczny jednocześnie poniesiemy.

Popyt warunkiem zrównoważonego rozwoju ADM

Skupienie na pracach badawczo-rozwojowych i wykorzystaniu gospodarczym systemów ADM prowadzi do strategii czysto podażowej, która zakłada, że Polska jest krajem, który – w najlepszym wypadku – dostarcza rozwiązań ADM wdrażanych w innych miejscach na świecie. Już dzisiaj polskie firmy z sektora AI w przeważającej mierze pracują dla klientów zagranicznych.

Tymczasem zrównoważony rozwój technologiczny wymaga również strategii popytowej, której celem powinno być zadbanie o to, byśmy jako społeczeństwo umieli wykorzystywać technologię świadomie i dbać dzięki niej o realizację własnych potrzeb. Spojrzenie na rozwój ADM i AI z perspektywy popytowej rodzi duże wyzwanie, ponieważ Polska od lat pozostaje jednym z państw europejskich o najniższym poziomie rozwoju cyfrowego. Pozostajemy w tyle zarówno pod względem rozwoju technologii cyfrowych oraz opartych na nich produktów i usług, jak i naszej zdolności ich wykorzystania przez jednostki, firmy, instytucje i organizacje. Samo zapóźnienie jest zrozumiałe, jeśli wziąć pod uwagę stan technologicznego zapóźnienia Polski odziedziczony w 1989 roku. Niestety, aktualna dynamika zmian rozwoju cyfrowego w Polsce nie jest na tyle wysoka, aby liczyć na znaczącą poprawę pozycji naszego kraju na tle innych państw członkowskich UE⁶.

Skupienie na społecznym wymiarze systemów ADM pozwala też przenieść dyskusję wyłącznie z obszaru podaży i uwzględnić także popyt na te rozwiązania. Jest to bowiem kwestia odpowiedniej świadomości społecznej dotyczącej oddziaływania technologii oraz zapewnienia kompetencji niezbędnych, by te technologie wykorzystać w najróżniejszych instytucjach i sferach życia.

⁶ The Digital Economy and Society Index (DESI) 2019, <https://ec.europa.eu/digital-single-market/en/desi>

Strategia popytowa, a nie tylko podażowa, jest niezbędna, by wykonać technologiczny żabi skok. Wymaga on bowiem społeczeństwa, instytucji i organizacji wykorzystujących technologie świadomie, nie tylko w celu osiągnięcia zysku ekonomicznego. Rozmawiając o zasadach regulacji algorytmów i tworząc ramy kontroli społecznej nad systemami ADM, możemy jednocześnie znacząco poszerzyć podstawową społeczną znajomość tematu, zwiększyć zaufanie do nowej technologii. W ten sposób powstać może fundament, na którym da się budować kolejne kompetencje wymagane do wdrożenia i przyswojenia systemów ADM.

Zaleta zapóźnienia: mądry Polak po czyjejsz szkodzie

Gdy obserwuje się rozwój systemów ADM na całym świecie, dostrzec można, że pierwsze rozwiązania są obarczone ryzykiem popełnienia błędów mających skutki społeczne. Rozwój technologiczny wyprzedza regulacje i inne formy nadzoru społecznego, systemy ADM niejednokrotnie powodują różne negatywne skutki społeczne – opisujemy je w rozdziałach drugim i trzecim. Złe opracowane zbiory danych lub błędy w rozumowaniu algorytmów są źródłem najróżniejszych skrzywień i nieodpowiednich decyzji. Brak przejrzystości systemów ADM uniemożliwia ocenę skutków ich oddziaływania.

Dzisiaj, gdy od kilku lat trwają zarówno powszechne wdrożenia tej technologii, jak i debata społeczna na ich temat, widać wyraźną zmianę w podejściu do systemów ADM. Coraz więcej uwagi poświęca się redukowaniu negatywnych skutków, takich jak skrzywienia w podejmowaniu decyzji. Debata regulacyjna coraz precyzyjniej opisuje narzędzia pozwalające nadać systemom ADM odpowiedni kształt. Podmioty polskie, wdrażając dziś systemy ADM, mogą więc to robić w dużo bardziej świadomy i bezpieczny sposób.

Najważniejszym przykładem zmieniającego się stosunku do systemów ADM jest ewolucja podejścia do metafory czarnej skrzynki. Przez długi czas przyjęło się myśleć, że systemy ADM są tzw. czarnymi skrzynkami: technologiami, które nie dają się zrozumieć, a więc też świadomie kontrolować. Dzisiaj podejście do tworzenia systemów ADM ulega dynamicznym zmianom. Jak stwierdziła fundacja Algorithm Watch w swoim manifestie, brak przejrzystości systemów ADM nie jest wynikiem oddziaływania praw natury. Od kilku lat coraz więcej dyskusji regulacyjnych zakłada, że czarne skrzynki dają się otworzyć. A więc zbiory danych mogą być weryfikowane, poprawiane i uzupełniane, co pozwala redukować skrzywienia i błędy w podejmowaniu decyzji. Wybory dokonywane przez algorytmy mogą być tłumaczone użytkownikom czy społeczeństwu i stać się przedmiotem potrzebnej debaty publicznej. Obywatele mają prawo i możliwości, by zrozumieć, dlaczego algorytm podjął określoną decyzję wpływającą na ich życie. Istnienie czarnych skrzynek ADM okazuje się być wadą kontroli społecznej nad systemami ADM, a nie cechą samej technologii.

Czarne skrzynki sztucznej inteligencji rodziły poczucie, że możemy jako ludzie utracić naszą autonomię, przekazując kluczowe decyzje w ręce automatycznych systemów. Narastająca dyskusja o systemach ADM i ich społecznym wymiarze pozwala odzyskać społeczny wymiar rozwoju technologicznego, budować poczucie społecznej kontroli i władzy nad technologią.

Polska droga do technologii w służbie społeczeństwa

Badania porównawcze pokazują, że w Polsce istnieje mniej wdrożeń systemów ADM niż w innych państwach europejskich i rozwiniętych gospodarkach⁷. Fakt ten należy traktować nie tylko jako oznakę cywilizacyjnego zapóźnienia, lecz także – a nawet przede wszystkim – jako szansę na zapewnienie zrównoważonego rozwoju ADM w Polsce. Wdrażając ADM, możemy bowiem wyciągać wnioski i uczyć się na błędach popełnianych w innych państwach.

Mniejsza liczba wdrożeń nie oznacza, że ADM jest nieobecny w naszym życiu. Na polskie społeczeństwo już dziś oddziałują systemy ADM – są to jednak zastosowania, nad którymi jako społeczeństwo mamy niewielką kontrolę. To przede wszystkim systemy ADM stosowane przez platformy internetowe – wyszukiwarki (np. Google), media społecznościowe (np. Facebook), ale też platformy *sharing economy* (np. Uber). Z rozwiązań ADM korzysta też większość banków, na przykład przy podejmowaniu decyzji kredytowych. Cechą tych systemów jest mała możliwość sprawowania nad nimi kontroli społecznej, co oznacza, że nie możemy w Polsce kształtować ich w sposób świadomy i zrównoważony.

Wykorzystanie szansy na żabi skok wymaga stworzenia przestrzeni debaty i wypracowywania społecznych norm kształtujących systemy ADM. To z kolei oznacza stworzenie odpowiedniego ekosystemu: podmiotów zdolnych do uczenia się na podstawie doświadczeń zagranicznych i świadomego kształtowania systemów zautomatyzowanego podejmowania decyzji. Dzięki temu polskie instytucje będą w stanie uczestniczyć także w debacie na poziomie europejskim i światowym – inaczej zasady nas obowiązujące będą ustalone bez nas. Wytworzenie odpowiednich instytucji zapewniających zrównoważony rozwój i nadzór nad ADM, co powinno nastąpić przed wdrożeniem tych rozwiązań, jest kluczem do wykorzystania renty zapóźnienia.

⁷ *Automating Society – Taking Stock of Automated Decision-Making in the EU*, <https://algorithmwatch.org/en/automating-society/>

Systemy ADM – przykłady z życia

Podczas analizy społecznych skutków wykorzystywania systemów ADM bez znaczenia jest ich klasyfikacja czy technologia wykorzystywana przez dany system. Kluczowy jest brak pełni ludzkiej kontroli nad procesem decyzyjnym wywierającym wpływ na jednostki i społeczeństwo. Dlatego w tym rozdziale prezentujemy wybór różnych zastosowań systemów ADM, bez podziału na ich rodzaje, ale w kluczowych dla życia społecznego obszarach: prewencji przestępstw, rynku pracy, opieki społecznej i medycznej, edukacji, dostępu do kultury i informacji, finansów i sądownictwa.

Pokazujemy zarówno systemy wykorzystywane przez podmioty publiczne, jak i przez firmy prywatne – naszym zdaniem również one mają znaczący wpływ na współczesne społeczeństwo. Zebraliśmy i przeanalizowaliśmy przykłady z Polski, Europy i ze świata.

Nasz katalog ma charakter przeglądowy. Nie jest kompletnym zestawem systemów ADM wykorzystywanych na całym świecie, lecz wyborem przykładów mających pomóc zrozumieć skalę stojącego przed nami wyzwania. Opisuując przykłady, staramy się też pokazać skutki społeczne ich wdrożenia, w tym największe wyzwania. W niektórych przypadkach pokazujemy proces naprawy tych systemów poprzez różnorodne działania zmierzające do redukcji szkód związanych z wprowadzaniem ADM.

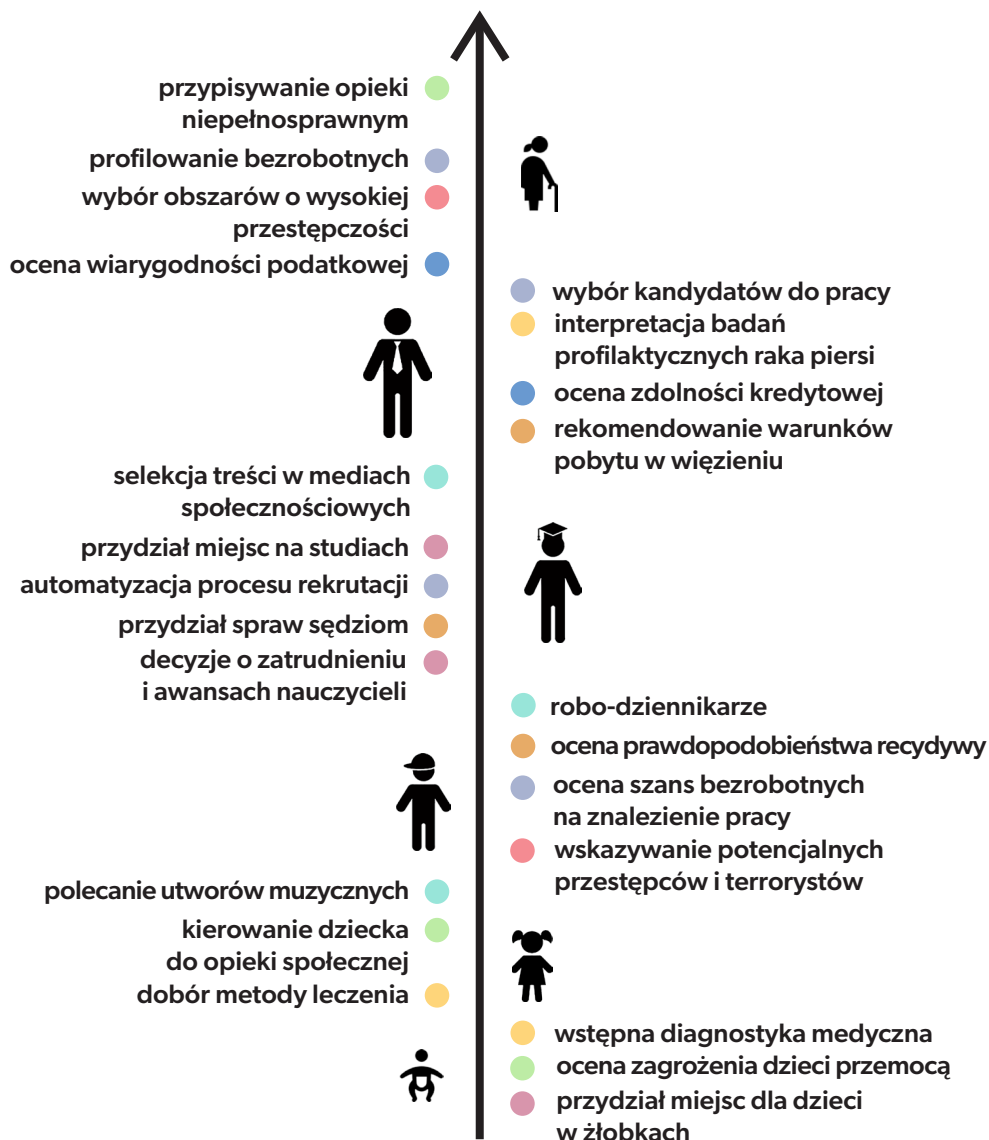
Charakterystyka systemów zautomatyzowanego podejmowania decyzji

Systemy ADM mają kilka istotnych cech:

- zazwyczaj wymagają dostępności dużych zbiorów danych (często osobowych), na bazie których podejmowane są decyzje;
- w różnym stopniu angażują człowieka w podejmowanie decyzji, w niektórych wypadkach ludzie są z niego wyłączeni i decyzja jest podejmowana całkowicie automatycznie; w innych ADM pełni funkcję sugerującą;
- w różnym stopniu wykorzystują proste algorytmy decyzyjne, metody statystyczne oraz technologie sztucznej inteligencji;
- są stosowane do automatyzacji procesów istotnych społecznie, w sferach takich jak służba zdrowia, sądownictwo, rynek pracy i zatrudnienie czy decyzje kredytowe (dlatego też nie są dla nas systemami ADM systemy wykorzystywane np. do automatyzacji pracy w fabrykach).

Algorytmy decydują za nas

Przykłady rozwiązań ze świata wykorzystujących zautomatyzowane podejmowanie decyzji w różnych obszarach życia



- | | | | |
|-------------------------|---------------|--------------------|----------------------------------|
| ● prewencja przestępstw | ● rynek pracy | ● opieka społeczna | ● opieka medyczna |
| ● edukacja | ● finanse | ● sądownictwo | ● dostęp do informacji i kultury |

Systemy zautomatyzowanego podejmowania decyzji możemy podzielić na trzy kategorie w zależności od zaangażowania człowieka w podejmowane decyzje⁸:

- w pełni autonomiczne – działające bez nadzoru człowieka;
- z człowiekiem w procesie (*human-in-the-loop*) – w których system podporządkowany jest temu, co człowiek nakazuje wykonać; człowiek bierze udział w procesie tworzenia decyzji;
- z człowiekiem „na” procesie (*human-on-the-loop*) – w których człowiek nadzoruje działalność zautomatyzowanego systemu – otrzymuje wynik, który ma prawo zmienić.

Poziom zaangażowania człowieka w podejmowanie decyzji jest z perspektywy społecznej istotniejszym czynnikiem niż poziom wykorzystania technologii maszynowego uczenia się. Zachowanie czynnika ludzkiego jest bowiem kluczowe do zapewnienia ochrony praw obywateli. Zapewnia bowiem przejrzystość i odpowiedzialność za proces, których nie potrafimy dziś zapewnić w przypadku systemów w pełni zautomatyzowanych. Na podstawie art. 22 Rozporządzenia o Ochronie Danych Osobowych (RODO) obywatel może zażądać, by w istotnych kwestiach – generujących skutki prawne – decyzja w oparciu o przetwarzanie danych osobowych nie była podejmowana w pełni automatycznie, by człowiek był obecny w pętli decyzyjnej. Pytania o zakres społecznej kontroli nad podejmowaną decyzją, jej racjonalność i ewentualne skrzywienia pojawiają się w odniesieniu do wszystkich trzech kategorii ADM. Należy jednak podkreślić, że coraz częściej systemy ADM, które miały mieć jedynie funkcję doradczą bądź sugerującą, w praktyce zastępują decyzje człowieka. Dzieje się tak, ponieważ człowiek podejmujący decyzje niemal bezkrytycznie ufa propozycji danego systemu.

Stosując pojęcie ADM, przenosimy nacisk debaty z poziomu samej technologii i zastosowanych rozwiązań na kwestie społeczne. Patrzymy na system społeczno-technologiczny, w którym działają różne technologie – zarówno te oparte na uczeniu maszynowym, jak i prostsze algorytmy. W dyskusjach o systemach ADM punkt ciężkości przenosi się z analizy strony technicznej (modelu podejmującego decyzję w oparciu o dane) na szerszy proces obejmujący także procesy tworzenia i programowania systemów ADM oraz nadzoru nad ich działaniem.

Prewencja przestępstw

Systemy zautomatyzowanego podejmowania decyzji są coraz chętniej wykorzystywane w prewencji przestępstw (*predictive policing*), głównie na poziomie analizy danych statystycznych związanych z przestępczością. Systemy ADM określają w ten sposób obszary i społeczności szczególnie zagrożone ryzykiem zwiększonej przestępczości, a także wskazują osoby, które z większym prawdopodobieństwem popełnią przestępstwo. Prewencja przestępstw jest jednym z bardziej kontrower-

⁸ *Human-in-the-counterfactual-loop*, <https://ai-alignment.com/counterfactual-human-in-the-loop-a7822e36f399>

syjnych obszarów stosowania ADM. Delegowanie kompetencji organów ścigania związanych z oceną zagrożenia polegającego na popełnieniu przestępstwa i powierzenie elementu oceny automatycznym systemom niesie ze sobą duże ryzyko ograniczenia praw i wolności jednostek. Systemy związane z ADM dotyczące prewencji przestępstw często bazują na tzw. *dirty data*⁹ – zbiorach danych, które są niepoprawnie skonstruowane, niekompletne lub niespójne. Badania pokazują¹⁰, że takie systemy potęgują zjawisko dyskryminacji. W typowym scenariuszu algorytm opiera się na historycznych danych, nie kwestionując ich. W ten sposób przeszłe uprzedzenia – dotyczące na przykład osób czarnoskórych – są powielane przez algorytm.

- ◆ – Polska
- – Europa
- ◇ – Świat

◇ Wykorzystanie historycznych danych statystycznych dotyczących przestępstw w celu identyfikacji obszarów, gdzie z większym prawdopodobieństwem wystąpią włamania do mieszkań czy kradzieże samochodów coraz częściej determinuje decyzje co do rozmieszczenia patroli policji. Jednak badania¹¹ przeprowadzone przy okazji wdrażania systemu Predictive Policing P4 w Stuttgarcie w latach 2015–2016 przez Instytut Maxa Plancka wykazały brak pozytywnych efektów takiego podejścia – nie powodowały zmniejszenia liczby przestępstw. Niemniej systemy oparte na idei *predictive policing* są wdrażane w coraz większej liczbie miejsc na świecie. Wymienić możemy tutaj system PredPol w Los Angeles¹² i system wykorzystywany w regionie Westkust w Belgii¹³ od 2016 roku.

► Innym pomysłem na wykorzystywanie systemów ADM w celu prewencji przestępstw jest analiza danych osobowych i wskazywanie osób, które zostają ocenione jako prawdopodobni przyszli sprawcy przestępstw. Takim systemem jest np. system Hessen-Data¹⁴ wdrożony przez rząd niemieckiej Hesji w 2017 roku. System ten (jak się podejrzewa, bowiem informacje o nim nie są dostępne) łączy dane z mediów społecznościowych z policyjnymi bazami danych oraz danymi o połączeniach telefonicznych w celu identyfikacji potencjalnych terrorystów. Oprogramowanie jest dostarczane przez Palantir, prywatną firmę programistyczną z USA.

► DANTE¹⁵ (akronim oznaczający „Wykrywanie i analizowanie treści internetowych i finansowania związanych z terroryzmem”) to eksperymentalny projekt finansowa-

⁹ With ‘dirty data’, predictive policing can make bias worse, <https://www.futurity.org/predictive-policing-data-2003012-2/>

¹⁰ Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3333423

¹¹ Predictive Policing: Evaluation des baden-württembergischen Pilotprojekts P4, Freiburg 2017, <https://www.mpicc.de/de/forschung/projekte/predictive-policing/>

¹² „Predictive policing”: Law enforcement revolution or just new spin on old biases? Depends who you ask, <https://www.cbc.ca/news/world/crime-los-angeles-predictive-policing-algorithms-1.4826030>

¹³ Politiezone Westkust experimenteert met datasets in strijd tegen criminaliteit, https://www.standaard.be/cnt/dmf20160517_02292901

¹⁴ Hessischer Landtag: Drucksache 19/6864, <http://starweb.hessen.de/cache/DRS/19/4/06864.pdf>

¹⁵ DANTE- Detecting and analysing terrorist-related online contents and financing activities, <https://www.h2020-dante.eu/>

nyprzez Komisję Europejską w ramach Horyzontu 2020, mający na celu wykorzystanie zautomatyzowanego podejmowania decyzji w celu zwalczania terroryzmu. W projekcie bierze udział 18 krajów, a DANTE ma zapewnić możliwość analizy danych, co ma umożliwić organom ścigania i agencjom wywiadowczym lepsze wykrywanie sieci terrorystycznych.

◇ Prewencja przestępstw w niektórych regionach świata związana jest z poważnymi naruszeniami praw człowieka. Zintegrowana platforma wspólnych operacji (IJOP, 一体化 联合 作战 平台) to jeden z głównych systemów stosowanych przez chińskie władze do masowego nadzoru w Sinkiang (Xinjiang)¹⁶. Platforma ta gromadzi dane o ludziach i automatycznie sugeruje urzędnikom, których obywateli uważa za potencjalnie groźnych – w efekcie wiele osób zostało zatrzymanych i osadzonych w obozach reedukacyjnych.

Pomysł na wykorzystywanie systemów automatycznego podejmowania decyzji w celu zapobiegania przestępstwom jest również bardzo nośny politycznie. Donald Trump zaproponował ostatnio¹⁷, żeby media społecznościowe miały za zadanie namierzać potencjalnych sprawców strzelanin.

Rynek pracy

Na rynku pracy systemy zautomatyzowanego podejmowania decyzji wykorzystywane są zarówno przez administrację publiczną w procesach dotyczących osób bezrobotnych, jak i przez firmy prywatne w rekrutacji. Ma to na celu przede wszystkim wstępną weryfikację kandydatów na dane stanowisko i szybką, automatyczną analizę danych o dużej ilości osób. Jednak nie tylko procesy rekrutacyjne zmienia się pod wpływem systemów ADM. Również sam rynek pracy ewoluuje pod ich wpływem – coraz częściej spotyka się argumenty, że w przypadku niektórych platform *sharing economy* to sam system ADM staje się pracodawcą.

◆ *W Polsce głośno stało się o wykorzystywaniu zautomatyzowanego podejmowania decyzji do profilowania bezrobotnych. Wszystkie osoby zarejestrowane jako bezrobotne dzielone były na trzy grupy (tzw. profile) różniące się stopniem gotowości do podjęcia pracy i rodzajem pomocy, którą otrzymają z urzędu pracy. Podział odbywa się na podstawie specjalnej tabeli: w trakcie rozmowy z urzędnikiem wypowiedzi bezrobotnego dopasowywane są do specjalnego klucza i punktowane, żeby ostatecznie stać się podstawą do przyznania bezrobotnemu określonego profilu. Ten bardzo prosty przykład (nawet nie algorytmu, a tabeli, która jednak dla większości urzędników była podstawą ostatecznej decyzji) pokazał, jakie problemy mogą powstać w związku z wykorzystywaniem ADM w profilaktyce bezrobocia: brak przejrzystości zautomatyzowanych procedur, brak odpowiedniej procedury odwoławczej oraz zagrożenie niesprawiedliwym*

¹⁶ China's Algorithms of Repression, <https://www.hrw.org/report/2019/05/01/chinas-algorithms-repression/reverse-engineering-xinjiang-police-mass-surveillance>

¹⁷ Trump wants social media to detect mass shooters before they commit crimes, <https://www.vox.com/recorde/2019/8/5/20754790/trump-social-media-detect-shooter-crime>

profilowaniem przez uproszczony algorytm nieuwzględniający złożonych sytuacji życiowych. Fundacja Panoptykon doprowadziła do uznania tego systemu za niekonstytucyjny przez Trybunał Konstytucyjny – jednak podstawą nie była szczegółowa analiza działania ADM, a jedynie wadliwy tryb zbierania danych o jednostkach¹⁸.

► W Austrii od 2018 roku Arbeitsmarktservice (instytucja odpowiedzialna za przeciwdziałanie bezrobociu) zaczęła automatycznie oceniać szanse osób bezrobotnych na znalezienie pracy w oparciu o modele statystyczne. Jak na razie nie wykorzystano tego systemu do przyznawania benefitów (ma się to jednak zmienić w 2020 roku)¹⁹.

◆ Na rynku dostępnych jest też coraz więcej systemów wspierających procesy zatrudnienia z wykorzystaniem ADM – tzw. *robo-recruiting*²⁰. Systemy te przeszukują profile kandydatów pod kątem określonych kwalifikacji i słów kluczowych, zazwyczaj na wstępnych etapach selekcji.

◇ W 2014 roku firma Amazon zaczęła pracować nad własnym systemem automatyzacji rekrutacji, opartym na archiwalnych danych rekrutacyjnych. System porównywał dane zbierane podczas rekrutacji z późniejszymi wynikami pracowników. Dane, które stanowiły podstawę szkolenia algorytmu, pochodziły z wcześniejszych praktyk rekrutacyjnych firmy. Z czasem okazało się, że w tym czasie rekrutowano głównie mężczyzn na stanowiska menedżerskie²¹. Na skutek tego algorytm rekrutacyjny bazował na tym dyskryminującym założeniu. Amazon po trzech latach zrezygnował z tworzenia systemu.

◇ Alex Rosenblat w swojej książce *Uberland* prezentuje tezę²², że na platformach internetowych takich jak Uber to algorytmy stają się pracodawcami. W takich systemach ADM algorytm pełni decydującą rolę w wyborze osoby, której przydzielane jest zadanie do wykonania.

Opieka społeczna

Kluczowym elementem systemu opieki społecznej jest podejmowanie decyzji o przyznaniu świadczeń konkretnym osobom. Zazwyczaj przesłanki przyznania danego świadczenia są kwantyfikowalne, przez co coraz chętniej wprowadza się systemy ADM do opieki społecznej. Stosowane są one na różnych etapach: od identyfikacji

¹⁸ Profilowanie bezrobotnych niezgodne z Konstytucją, <https://panoptykon.org/wiadomosc/profilowanie-bez-robotnych-niezgodne-z-konstytucja-0>

¹⁹ AMS bewertet Arbeitslose künftig per Algorithmus, <https://www.derstandard.at/story/2000089095393/ams-bewertet-arbeitslose-kuenftig-per-algorithmus>

²⁰ Robo-recruiting: How soon is it before the machines take over?, <https://www.executivegrapevine.com/publications/magazine/august-2018/robo-recruiting-how-soon-will-you-be-replaced-by-a-machine>

²¹ Amazon scraps secret AI recruiting tool that showed bias against women, <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>

²² We've Been Trapped in 'Uberland', <https://www.citylab.com/life/2018/10/uberland-and-how-ride-hailing-changed-work/573586/>

osób zagrożonych pewnymi zjawiskami społecznymi, po przyznawanie konkretnych godzin opieki poszczególnym osobom. Jednak bardzo często te pomysły spotykają się z dużą krytyką społeczną²³ ze względu na ochronę prawa do prywatności oraz to, że systemy ADM często powielają funkcjonujące społecznie stereotypy.

► Duńska gmina Gladsaxe planowała wykorzystywać dane pochodzące z różnych rejestrów w celu identyfikowania dzieci zagrożonych przemocą lub zaniedbaniem²⁴. Dane z obserwacji dzieci w instytucjach edukacyjnych miały być łączone z danymi o rodzicach (m.in. dane zdrowotne, o przestępstwach czy podatkowe). Krytycy podnosili wyzwania związane ze sprawiedliwym skonstruowaniem algorytmu podejmującego decyzje o kwestiach wrażliwych społecznie oraz sprawę zagrożenia nadmiernego nadzoru nad dziećmi i ich rodzinami. Pod wpływem krytyki wdrożenie systemu zostało zawieszone.

◇ Systemy, które w Europie spotykają się z dużym sprzeciwem społecznym, są od kilku lat wdrażane w Stanach Zjednoczonych. W 2013 roku instytucje opieki społecznej w Los Angeles zaczęły stosować algorytmy w celu oceny zagrożenia dzieci i potrzeby skierowania ich do systemu opieki społecznej. Komercyjny system przygotowany przez firmę SAS był od początku krytykowany za brak przejrzystości, a przeprowadzone audyty wykazały bardzo wysoki poziom wyników fałszywie pozytywnych: system wskazywał jako zagrożone rodziny, które w rzeczywistości nie potrzebowały pomocy. W ewaluacji projektu²⁵ podkreślono, że system nie przedstawiał kompleksowego obrazu danej rodziny, ale zamiast tego skupiał się na kilku przypadkowo wybranych cechach, nadając im nadmierną wagę.

◇ *Również w Stanach Zjednoczonych, w stanie Arkansas, zaczęto używać zautomatyzowanego systemu podejmowania decyzji do przypisywania godzin pielęgnarskich przysługujących danej osobie z niepełnosprawnością²⁶. Algorytm przypisał większości korzystających ze wsparcia znacznie mniejszą liczbę godzin niż ta, która przysługiwała im wcześniej w wyniku oceny przez eksperta. Dodatkowo osoby z niepełnosprawnością były pozbawione mechanizmu odwoływania się od decyzji. Legal Aid pozwała stan Arkansas, a sąd orzekł, że algorytm jest błędny i niekonstytucyjny.*

Opieka medyczna

Zautomatyzowane podejmowanie decyzji, w postaci systemów sugerujących, coraz częściej wykorzystywane jest w medycynie. W dziedzinie tej najwięcej systemów

²³ Councils use 377,000 people's data in efforts to predict child abuse, <https://www.theguardian.com/society/2018/sep/16/councils-use-377000-peoples-data-in-efforts-to-predict-child-abuse>

²⁴ Regeringen vil overvåge alle landets børnefamilier og uddele point, <https://politiken.dk/indland/art6365403/Regeringen-vil-overv%C3%A5ge-alle-landets-b%C3%B8rnefamilier-og-uddele-point>

²⁵ Examination of Using Structured Decision Making and Predictive Analytics in Assessing Safety and Risk in Child Welfare, http://file.lacounty.gov/SDSInter/bos/bc/1023048_05.04.17OCPRReportonRiskAssessmentTools_SDMandPredictiveAnalytics_.pdf

²⁶ A Healthcare Algorithm Started Cutting Care, and No One Knew Why, <https://www.theverge.com/2018/3/21/17144260/healthcare-medicaid-algorithm-arkansas-cerebral-palsy>

ADM korzysta ze sztucznej inteligencji i analizy *big data*. Są one wykorzystywane w diagnostyce chorób, leczeniu stanów przewlekłych, zarządzaniu systemami usług zdrowotnych, badaniami nad lekami²⁷. Sztuczna inteligencja i pośrednio również systemy ADM mogą pomóc w rozwiązywaniu wyzwań medycznych, jednak dużym ograniczeniem jest dostępność i jakość danych niezbędnych do prawidłowego funkcjonowania sztucznej inteligencji.

◆ Wiele z produktów medycznych w tym zakresie tworzonych jest w Polsce (co nie oznacza, że są tutaj wykorzystywane). Ardigen S.A., bioinformatyczna spółka-córka giełdowej Selvity, podjęła współpracę z EpiBiome, amerykańską firmą specjalizującą się w precyzyjnej inżynierii mikrobiomu, rozwijającą terapie medyczne będące alternatywą dla leczenia antybiotykami. Zakres współpracy koncentruje się na pracach nad algorytmami i bazami danych służącymi do skuteczniejszej identyfikacji bakterii. Dzięki czemu analizy mają być poprawniejsze i możliwe będzie stosowanie alternatywnych terapii medycznych²⁸.

◆ Infermedica, polska firma, która wykorzystuje sztuczną inteligencję do wstępnej diagnostyki medycznej, pozyskała ostatnio 14 mln zł (3,65 mln dolarów) w rundzie *seed*, której przewodzi estoński fundusz Karma Ventures. Firma przeznaczy nowe środki na zwiększenie swojej obecności w Stanach Zjednoczonych i Europie oraz przyspieszenie rozwoju technologii²⁹.

► W 2017 roku region stołeczny Danii zawarł umowę z IBM³⁰ w celu wykorzystywania systemu Watson do rutynowych badań profilaktycznych raka piersi. Argumentem przemawiającym za taką decyzją był brak lekarzy specjalizujących się w radiologii. Decyzja ta wywołała jednak duże kontrowersje – umowę zawarto pomimo doniesień o trudnościach podczas poprzedniego testu z innym narzędziem Watson Oncology. Jedynie w 27% przypadków rekomendacje Watson Oncology zgadzały się z decyzjami lekarzy dotyczącymi leczenia³¹. Najprawdopodobniej było to spowodowane przeszkoleniem systemu w Stanach Zjednoczonych według lokalnych wytycznych medycznych. Przy okazji tych badań naukowcy z Uniwersytetu IT w Kopenhadze wezwali do informowania opinii publicznej i decydentów o możliwych błędach algorytmów w medycynie³².

²⁷ Bioethics briefing note: Artificial intelligence (AI) in healthcare and research, <http://nuffieldbioethics.org/wp-content/uploads/Artificial-Intelligence-AI-in-healthcare-and-research.pdf>

²⁸ Ardigen dostarcza rozwiązania bioinformatyczne dla wschodzącej gwiazdy rynku biotech z San Francisco, <https://www.kierunekfarmacja.pl/artukul,27747,ardigen-dostarcza-rozwiazania-bioinformatyczne-dla-wschodzacej-gwiazdy-rynku-biotech-z-san-francisco.html>

²⁹ Infermedica pozyskała 14 mln zł na rozwój platformy do wstępnej diagnostyki medycznej, <https://www.wroclaw.pl/startupy/infermedica-pozyskala-14-mln-zl-na-rozwoj-platformy-do-wstepnej-diagnostyki-medycznej>

³⁰ Ny aftale om udvikling af kunstig intelligens i sundhedsvæsenet, <https://www.regionh.dk/presse-og-nyt/pressemeddelelser-og-nyheder/Sider/Ny-aftale-om-udvikling-af-kunstig-intelligens-i-sundhedsv%C3%A6senet.aspx>

³¹ Doktor Watson i modvind: Foreslog livsfarlig medicin til danske patienter, <https://ing.dk/artikel/doktor-watson-modvind-foreslog-livsfarlig-medicin-danske-patienter-207529>

³² Lad os ikke forføre af vidunderrobotten Watson, <https://finans.dk/debat/ECE9742204/lad-os-ikke-forfoere-af-vidunderrobotten-watson/?ctxref=ext>

► Wiele projektów medycznych związanych z algorytmami rozwijanych jest również na uczelniach. Przykładowo stworzony przez Università Cattolica del Sacro Cuore w Mediolanie projekt „Abbiamo i numeri giusti” („Mamy odpowiednie liczby”) to zautomatyzowany system pomagający instytucjom służby zdrowia dobrać najlepsze leczenie dla danego przypadku, jednocześnie optymalizując wydatki publiczne na służbę zdrowia³³.

Edukacja

Algorytmy są coraz częściej wykorzystywane również do podejmowania decyzji dotyczących edukacji od przypisywania uczniów do placówek zaczynając, poprzez delegowanie nauczycieli do miejsca pracy, na ewaluacji nauczycieli kończąc.

◆ We Wrocławiu głośno było w 2018 roku o systemie przypisywania dzieci do żłobków, który okazał się porażką ze względu na zastosowany algorytm³⁴. Do żłobków dostały się dzieci, które nie powinny, zaś te, którym miejsce powinno przysługiwać, się nie dostały. Ze względu na brak informacji o działaniu systemu rodzice zarzucali mu nieprzejrzyste i krzywdzące podejmowanie decyzji w sytuacji, gdy liczbą żłobków była niewystarczająca. W Polsce nieznane są całościowe dane o wykorzystaniu systemów ADM w alokacji dzieci i młodzieży do placówek oświatowych. Platforma Zarządzania Oświatą³⁵ stworzona przez Asseco Data Systems jest wykorzystywana przez 4 500 szkół w 20 miastach, jednak nie wiadomo, na ile Platforma wykorzystuje systemy ADM.

► W 2018 roku rząd francuski przyspieszył reformę, zgodnie z którą uniwersytety zaczęły mieć prawo do odrzucenia wniosków o przyjęcie od potencjalnych studentów. Reforma została wprowadzona wraz z uruchomieniem internetowego narzędzia o nazwie Parcoursup, które przypisywało studentów do wolnych miejsc na uczelniach. Zadaniem platformy było łączenie życzenia studentów co do wymarzonej uczelni z dostępnymi ofertami. Parcoursup jest krytykowany za nieprzejrzystość procesu decyzyjnego – użytkownicy systemu nie są informowani, że decyzje są podejmowane automatycznie, a rząd opublikował jedynie kod samej platformy, a nie algorytmu. Uważa się, że system wykorzystuje dane osobowe, aby pomóc uniwersytetom w wyborze studentów, którzy najlepiej pasują do określonych kierunków studiów³⁶.

► We Włoszech w ramach reform edukacyjnych „Buona Scuola” wprowadzono zautomatyzowany system zarządzania mobilnością wykładowców akademickich między uniwersytetami w roku akademickim 2016/17. System miał traktować priorytetowo

³³ *L'algoritmo che fa bene alla salute*, <https://www.cattolicanews.it/l-algoritmo-che-fa-bene-alla-salute>

³⁴ *Zamieszanie z rekrutacją do żłobków. 299 dzieci najpierw się dostało, później je wykreślili*, <https://wroclaw.wyborcza.pl/wroclaw/7,35771,23544016,zamieszanie-z-rekrutacja-do-zlobkow-299-dzieci-najpierw-sie.html?disableRedirects=true>

³⁵ *Platforma Zarządzania Oświatą, Asseco Data Systems*, <https://www.assecods.pl/oferta/administracja-terenowa/platforma-zarzadzania-oswiata/>

³⁶ *Le gouvernement ouvre le code source de Parcoursup, mais pas ses algorithmes «locaux»*, <https://www.nextinpact.com/news/106617-le-gouvernement-ouvre-code-source-parcoursup-mais-pas-ses-algorithmes-locaux.htm>

wyniki oceny nauczycieli – jeżeli nauczyciel miał wyższy wynik, to jego preferencje dotyczące wolnych miejsc pracy w konkretnej lokalizacji miały być traktowane priorytetowo. Zamiast tego algorytm dawał pierwszeństwo jedynie preferencjom nauczycieli, bez uwzględnienia wyników. Wielu nauczycieli przypisanych zostało do pracy w miejscach, które nie były nawet na ich liście. W efekcie przeprowadzono kontrolę algorytmu na uniwersytetach La Sapienza oraz Tor Vergata. Autorzy raportu z kontroli stwierdzili³⁷, że algorytm ten nie powinien był łączyć nauczycieli z ich docelowymi miejscami pracy, bo był po prostu źle napisany i niedziałający.

◇ *Federacja Nauczycieli w Houston pozwała władze lokalne za wykorzystywanie danych z testów uczniów w automatycznym systemie podejmowania decyzji o zatrudnieniu nauczycieli i ich awansie³⁸. Władze lokalne twierdziły, że nie są w stanie wyjaśnić decyzji algorytmu kwestionującym ją nauczycielom, bo algorytm działa na zasadzie czarnej skrzynki (black box). Nauczyciele wnieśli więc pozew, zarzucając naruszenie przepisów konstytucyjnych, praw obywatelskich i prawa pracy, a do stawka systemu nawet w trakcie procesu odmawiał zaprezentowania algorytmu. W efekcie sąd orzekł, że korzystanie z takich systemów w sferze publicznej narusza prawo do sprawiedliwego procesu i wyjaśnionej decyzji.*

Dostęp do informacji i kultury

Serwisy internetowe są, szczególnie w młodszych pokoleniach, głównym źródłem informacji, wiedzy i treści kulturowych. Serwisy społecznościowe i sieciowi agregatorzy treści wypierają źródła tradycyjne, takie jak telewizja. Jednocześnie serwisy te stosują algorytmy, które personalizują ofertę dla każdego użytkownika.

Symbolem takiego podejścia może być serwis Spotify, który na początku tygodnia przedstawia każdemu użytkownikowi spersonalizowaną playlistę na dany tydzień, wygenerowaną przez algorytm. Drugim przykładem jest Facebook, którego algorytm selekcjonuje treści widoczne dla danego użytkownika.

Jakie są tego skutki? Z jednej strony selekcja treści wydaje się jedynym sposobem radzenia sobie z natłokiem informacji. Z drugiej jednak – skutkuje to zamykaniem użytkowników w bańkach informacyjnych, możliwością skrzywionego obrazu świata czy kontrolą algorytmów polecających treści nad tym, co stanowi kanon kultury.

◇ Najbardziej kontrowersyjnym przykładem, który wpływa na dostęp do informacji i kultury milionów użytkowników, jest algorytm Facebooka odpowiadający za pozycjonowanie treści na tablicy (wall) danego użytkownika. W związku z krytyką tego mechanizmu Facebook w maju 2019 roku ogłosił, że nowy algorytm pozycjonujący został zmieniony w taki sposób, żeby przede wszystkim uwzględniać informacje

³⁷ Perizia tecnica preliminare sull'analisi dell'algoritmo che gestisce il software della mobilitadocenti per L'A.S. 2016/2017, <https://www.orizzontescuola.it/wp-content/uploads/2017/06/Perizia-tecnica-preliminare-pdf.pdf>

³⁸ *Litigating Algorithms: challenging government use of algorithmic decision systems*, <https://ainowinstitute.org/litigatingalgorithms.pdf>

o tym, kto z naszej listy znajomych jest dla nas najważniejszy, a także to, które informacje mogą mieć dla nas największą wartość³⁹. Krytycy takiego podejścia twierdzą, że doprowadzi to do promowania informacji i treści, które sprawiają, że użytkownicy są z siebie zadowoleni. Nadal nie rozwiązano jednak największego problemu – braku transparentności algorytmu. Nie wiadomo, jakie dane wejściowe wykorzystywane są do selekcji treści pokazywanych użytkownikom i tak naprawdę sami użytkownicy mają znikomą kontrolę nad zakresem baniek informacyjnych, w których są zamykani.

◇ Przykładem automatycznego podejmowania decyzji w zakresie tego, jakie utwory są widoczne dla użytkowników, jest funkcjonowanie serwisów muzycznych, jak np. Spotify. Spotify rekomenduje swoim użytkownikom nowe utwory i cotygodniowe playlisty stworzone specjalnie dla nich. Autorzy książki *Spotify Teardown: Inside the Black Box of Streaming Music*⁴⁰ zwracają jednak uwagę, jak duży wpływ na całą branżę mają algorytmiczne rekomendacje. Model biznesowy Spotify nie polega na pomaganiu słuchaczom w odkrywaniu nowej muzyki, ale na gromadzeniu informacji o słuchaczach w celu sprzedaży odbiorców reklamodawcom. Autorzy wskazują na sposób, w jaki Spotify zmienia sposób funkcjonowania serwisu na przestrzeni lat – konsumpcja muzyki w oparciu o wyszukiwanie została zastąpiona rekomendacjami wokół „zachowań, odczuć i nastrojów” użytkowników, analizowanych na podstawie wcześniejszych muzycznych wyborów⁴¹.

► Jednakże algorytmy nie są wykorzystywane tylko w selekcji treści dostępnych dla użytkowników, ale też coraz częściej w ich tworzeniu. Norweska Agencja Informacyjna od 2016 roku wykorzystuje dziennikarzy-roboty do przygotowywania informacji sportowych⁴². Dzięki temu raporty z meczów są już dostępne w 30 sekund po zakończeniu gry. Zautomatyzowane raporty informacyjne są nadzorowane przez zespół dziennikarzy i dotychczas okazały się w 99 procentach niezawodne. Jest to kolejny etap automatyzacji dziennikarstwa⁴³, który ma wpływ zarówno na jakość dziennikarstwa, jak i na stabilność zawodu dziennikarza.

◇ Ciekawym przykładem podejmowania decyzji w kulturze z wykorzystaniem analizy danych jest również tworzenie własnych treści przez platformy na podstawie danych o zachowaniach użytkowników. Przykładem jest serial *House of Cards* – Netflix zlecił produkcję 26 odcinków przed emisją pierwszego z nich, był bowiem przekonany o sukcesie komercyjnym produkcji⁴⁴. Pewność była skutkiem analizy danych o preferencjach użytkowników Netflixu – firma była w stanie ustalić, że fani oryginalnego

³⁹ Facebook changes algorithm to promote worthwhile & close friend content, <https://techcrunch.com/2019/05/16/facebook-algorithm-links/>

⁴⁰ Spotify Teardown: Inside the Black Box of Streaming Music, <https://mitpress.mit.edu/books/spotify-teardown>

⁴¹ How Spotify's algorithms are running music, <https://www.ft.com/content/dca07c32-6844-11e9-b809-6f0d2f5705f6>

⁴² Norwegian News Agency is betting on automation for football coverage, <https://www.journalism.co.uk/news/norwegian-news-agency-is-betting-on-automation-for-football-coverage/s2/a647189/>

⁴³ How artificial intelligence is transforming journalism, https://www.equaltimes.org/how-artificial-intelligence-is?utm_campaign=algoaware&lang=en#.XYbBcJMzaAz

⁴⁴ How Netflix built a House of Cards with big data, <https://www.idginsiderpro.com/article/3207670/how-netflix-built-a-house-of-cards-with-big-data.html>

House of Cards (emitowanego w Wielkiej Brytanii) oglądali również filmy z udziałem Kevina Spacey i wyreżyserowane przez Davida Finchera, jednego z producentów nowego serialu.

Finanse

Systemy ADM wykorzystywane są w finansach zarówno przez administrację publiczną, jak i banki. W pierwszym przypadku analiza danych służyć ma głównie uszczelnieniu systemu podatkowego i przeciwdziałaniu potencjalnym nadużyciom podatkowym, w drugim głównie lepszemu ustalaniu wiarygodności kredytowej.

◆ W 2017 roku wprowadzono zmiany ordynacji podatkowej⁴⁵ mające w założeniu służyć zwalczaniu nadużyć VAT-owskich. STIR jest systemem, który umożliwia wymianę informacji pomiędzy m.in. bankami a Krajową Administracją Skarbową. W STIR ustalany jest tzw. wskaźnik ryzyka (tj. wskaźnik wykorzystywania działalności banków i spółdzielczych kas oszczędnościowo-kredytowych do celów mających związek z wyłudzeniami skarbowymi⁴⁶). Ustalenie wskaźnika ryzyka, dokonywane przez algorytm, jest o tyle istotne, że jest on główną przesłanką żądania przez Szefa Krajowej Administracji Skarbowej blokady rachunku bankowego. Ustawa wprowadzająca system wyłącza dostęp do algorytmu ze względów bezpieczeństwa. W grudniu 2018 roku Sąd Administracyjny w Warszawie wydał precedensowy wyrok w sprawie STIR⁴⁷. Sąd stwierdził, że jeśli blokada konta została przedłużona o trzy miesiące, to nie musi być prowadzone postępowanie dowodowe – wystarczy, że po analizie przepływów pieniężnych urzędnik zakłada, iż istnieje ryzyko korzystania z banku w celu rozliczeń wyłudzeń podatkowych.

◆ Algorytmy wykorzystywane są przy podejmowaniu decyzji kredytowych czy pożyczkowych przez banki, jednak mechanizmy te chronione są tajemnicą działalności danego banku. Dotychczas klient nie znał przyczyn decyzji odmownej, czyli tak naprawdę nie wiedział, jakie dane okazały się decydujące w jego przypadku. Dzięki nowemu prawu do wyjaśnienia oceny zdolności kredytowej⁴⁸ każdy klient będzie mógł się dowiedzieć, jakie dane miały wpływ na decyzję w jego sprawie. Zmiana w prawie bankowym to element wdrażania do polskich przepisów ogólnego rozporządzenia o ochronie danych osobowych (RODO). Polska jako pierwsza wprowadziła prawo do wyjaśnienia decyzji dla każdej osoby ubiegającej się o kredyt.

► We Włoszech na przełomie 2016 i 2017 roku wprowadzono nowy fiskalny indeks wiarygodności podatkowej (*Indice Sintetico di Affidabilità*, ISA). ISA⁴⁹ jest średnią

⁴⁵ Dziennik Ustaw z 2017 r. poz. 2491, Ustawa z dnia 24 listopada 2017 r. o zmianie niektórych ustaw w celu przeciwdziałania wykorzystywaniu sektora finansowego do wyłudzeń skarbowych, www.dziennikustaw.gov.pl/du/2017/2491/D2017000249101.pdf

⁴⁶ STIR, czyli fiskus wszechwiedzący. Po co skarbowce wgląd w konto przedsiębiorcy?, <https://podatki.gazeta-prawna.pl/artykuly/1105020,stir-uszczelnianie-podatkow-wglad-w-konto-przedsiębiorcy.html>

⁴⁷ III SA/Wa 2057/18 - Wyrok WSA w Warszawie, <http://orzeczenia.nsa.gov.pl/doc/AB2D021C15>

⁴⁸ Fundacja Panoptykon: banki i finanse, <https://panoptykon.org/tematyczne/banki-i-finance>

⁴⁹ *Indice sintetico di affidabilità*, https://www.sose.it/sites/default/files/inline-files/Indici%20sintetici%20di%20affidabilita%CC%80_19.09.18.pdf

arytmetyczną zestawu podstawowych wskaźników wiarygodności fiskalnej, w tym spójności z poprzednimi deklaracjami fiskalnymi, konsekwentnej zgodności analizy przeszłych i obecnych anomalii na przestrzeni 8 lat. W wyniku analizy podmiotom przypisywany jest *scoring* fiskalny, a ideą ISA jest zrewolucjonizowanie logiki zgodności podatkowej. Ma to zakończyć epokę podejścia ukierunkowanego na kary i doprowadzić do stanu, w którym podmioty konsekwentnie wykazujące się wiarygodnością fiskalną będą nagradzane. W rezultacie podatnicy z wyższymi wynikami *scoringu* mogą cieszyć się systemem nagród obejmującym m.in. wyłączenie z niektórych kontroli skarbowych.

◇ Coraz większa liczba firm oferuje usługi wiarygodności kredytowej, z których korzystają różnego typu podmioty (ubezpieczyciele, dostawcy energii czy usług internetowych) w celu oceny zdolności kredytowej swoich potencjalnych klientów. Firmy oferujące taką usługę na masową skalę gromadzą informacje i dane o osobach pochodzące z różnorodnych źródeł. Czynią to w celu dokonania oceny, czy dana osoba jest wypłacalna, zazwyczaj wykorzystując do tego systemy ADM. Następnie firma ubezpieczeniowa, również często z wykorzystaniem systemów ADM, na podstawie tej oceny dokonuje decyzji, czy przyznać polisę danemu klientowi. Jest to jeden z przykładów sytuacji, w której systemy ADM mają wpływ na życie jednostek na masową skalę.

Sądownictwo

Wykorzystywanie systemów zautomatyzowanego podejmowania decyzji może mieć miejsce na wielu etapach: przypisywania spraw sędziom, zarządzania pracą sądów, ale też podejmowania decyzji o wymiarze kary.

ADM staje się w niektórych jurysdykcjach powszechnie stosowanym narzędziem (które postrzegane jest jako możliwość usprawnienia wymiaru sprawiedliwości^{50,51}) stanowiąc znaczące wyzwanie dla tradycyjnego modelu praw człowieka, regulacji prawnych⁵² i zasady sprawiedliwego procesu⁵³.

◆ W Polsce od początku 2018 roku funkcjonuje system Losowego Przydziału Spraw – to algorytm określa, który sędzia otrzyma konkretną sprawę do rozpatrzenia. Zdaniem Ministerstwa Sprawiedliwości ma to zapewnić bezstronność i transparentność w przypisywaniu danej sprawy do konkretnego sędziego. System ten wprowadzono we wszystkich 374 sądach powszechnych w całym kraju. Istnieją raporty, jak ten z kontroli w Sądzie Wojewódzkim w Toruniu, stwierdzające, że system nie działa poprawnie. Niektórzy sędziowie są przeciążeni pracą, inni nie są wybierani przez długi czas. W ciągu 7 miesięcy funkcjonowania systemu nierówność obciążenia po-

⁵⁰ *Even Imperfect Algorithms Can Improve the Criminal Justice System*, <https://www.nytimes.com/2017/12/20/upshot/algorithms-bail-criminal-justice-system.html>

⁵¹ *How AI Could Help the Public Sector*, <https://hbr.org/2018/01/how-ai-could-help-the-public-sector>

⁵² *How Big Data Is 'Automating Inequality'*, <https://www.nytimes.com/2018/05/04/books/review/automating-inequality-virginia-eubanks.html>

⁵³ *Artificial intelligence is not smart enough to set bail*, <https://www.sfoxaminer.com/opinion/artificial-intelligence-is-not-smart-enough-to-set-bail/>

szczególnych sędziów była wyraźnie zauważalna – niektórzy sędziowie otrzymywali o 5–8 spraw więcej niż inni. Fundacja ePaństwo chciała uzyskać dostęp do kodu źródłowego systemu i w grudniu 2017 roku złożyła do ministra wniosek o dostęp do informacji publicznej. Minister Sprawiedliwości odmówił udzielenia informacji wskazanej we wniosku, wskazując, że algorytm to informacja techniczna niebędąca informacją publiczną w rozumieniu polskiego prawa. Obecnie sprawa jest na etapie skargi kasacyjnej do Naczelnego Sądu Administracyjnego, wniesionej 14 grudnia 2018 roku. Podobne systemy funkcjonują też w Serbii, Słowacji i Gruzji⁵⁴.

◇ Coraz więcej stanów w USA wdraża przedprocesową ocenę ryzyka w oparciu o algorytmiczne instrumenty⁵⁵ wykorzystujące dane do oszacowania, czy dana osoba pojawi się na rozprawie i czy pozostając na wolności, będzie stanowiła zagrożenie dla porządku publicznego. Systemy te są przedstawiane jako sprawiedliwa alternatywa do obecnego systemu zabezpieczenia kaucją na etapie podejmowania decyzji o tymczasowym aresztowaniu. Jednak wiele środowisk zwraca uwagę, że algorytmy te są tak inteligentne, jak ich dane wejściowe. Wiele projektów przedprocesowego oceniania ryzyka korzysta z danych, które potwierdzają pewne społeczne uprzedzenia, np. to, że osoby rasy niekaukaskiej powinny częściej podlegać tymczasowemu aresztowaniu. Używanie takich danych do szkolenia systemów zautomatyzowanego podejmowania decyzji w systemie karnym jedynie utrwala tę korelację.

◇ Opracowany przez prywatną firmę Equivant (wcześniej Northpointe), COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*) ma przewidywać ryzyko popełnienia przez przestępcę innego przestępstwa. Działa poprzez zastrzeżony algorytm, który uwzględnia niektóre odpowiedzi na kwestionariusz zawierający 137 pytań. COMPAS jest wykorzystywany w niektórych stanach USA przez sędziów do decydowania o wymiarze kary nie tylko na podstawie popełnionego przestępstwa, ale też w oparciu o wynik z systemu. Eric Loomis, jeden ze skazanych, których wyrok został wydany w oparciu również o COMPAS, próbował podważyć legalność tej procedury. Sąd Najwyższy w Wisconsin oddalił jednak apelację⁵⁶, twierdząc, że wyrok byłby taki sam bez wykorzystania systemu. Podsumowując – w Stanach Zjednoczonych nadal algorytmy są używane do przygotowywania rekomendacji dotyczących wyroków w sprawach karnych oraz tego, czy jest szansa, że dana osoba popełni przestępstwo ponownie. Natomiast w Kanadzie systemy ADM wykorzystywane są do rekomendowania warunków osadzenia więźniów, w tym wyboru współwięźniów⁵⁷.

⁵⁴ alGOvrithms - State of play: Report on Algorithms Usage in Government-Citizens Relations in Czechia, Georgia, Hungary, Poland, Serbia and Slovakia, <https://epf.org.pl/en/wp-content/uploads/sites/3/2019/05/alGO-Vrithms-State-of-Play-Report.pdf>

⁵⁵ The use of pretrial "risk assessment" instruments: a shared state of civil rights concerns, <http://civilrightsdocs.info/pdf/criminal-justice/Pretrial-Risk-Assessment-Full.pdf>

⁵⁶ Supreme Court of Wisconsin: Wisconsin v. Loomis (881 N.W.2d 749), <https://www.scotusblog.com/wp-content/uploads/2017/02/16-6387-op-bel-wis.pdf>

⁵⁷ Supreme Court of Canada: Ewert v. Canada (2018 SCC 30), <https://scc-csc.lexum.com/scc-csc/scc-csc/en/item/17133/index.do>

Wyzwania i szanse związane z ADM

W niniejszym rozdziale przedstawiamy kluczowe wątki dyskusji o systemach ADM. Nie jest to lista kompletna – nie zawiera wszystkich padających w debacie argumentów. Stanowi jednak próbę podkreślenia szans i wyzwań stojących przed społeczeństwem we wdrażaniu takich systemów. Przedstawione wątki będą mieć różną wagę w każdym konkretnym przypadku. Już opisane w poprzednim rozdziale przykłady zastosowania ADM pokazują różnice w skali szans i społecznych wyzwań w zależności od różnych zastosowań systemu zautomatyzowanego podejmowania decyzji.

Piszemy o szansach i wyzwaniach, bo tworząc systemy ADM, trzeba przede wszystkim ustalić bilans zysków i strat. Systemy ADM jako takie nie są zagrożeniem – traktujemy je jako ogromną szansę mogącą przynosić różnorakie korzyści. Pod warunkiem, że zminimalizujemy ryzyka i szkody.

Przedstawione wyzwania nie są nierozwiązywalne – wręcz przeciwnie, w części z nich w ciągu ostatnich miesięcy dokonano znaczącego postępu. Najlepszym przykładem jest kwestia przejrzystości systemów ADM. Jeszcze rok temu algorytmy były opisywane jako czarne skrzynki, w które nie ma wglądu. Dziś inżynierowie, badacze i działacze społeczni nie rozmawiają o tym, czy istnieje możliwość tworzenia przejrzystych algorytmów – tylko jak najlepiej to zrobić.

Szanse

Większa efektywność

Systemy automatyzujące podejmowanie decyzji mają potencjał podniesienia efektywności działania instytucji publicznych oraz firm dzięki m.in. sztucznej inteligencji potrafiącej przejąć choć część najbardziej żmudnych elementów pracy od człowieka. Dzięki temu obsługa klienta, produkcja albo rozwiązywanie problemów obywatela jest szybsze, zajmuje mniej czasu pracowników lub urzędników, a przez to – jest tańsze.

Zautomatyzowane podejmowanie decyzji jest od paru dekad obecne w naszym życiu⁵⁸, na przykład w postaci autopilotów w samolotach. Choć piloci nadal wykonują przytłaczającą większość lądowań manualnie, większość czasu spędzonego w powietrzu samolot znajduje się pod kierownictwem oprogramowania, oczywiście na

⁵⁸ A framework for management information systems, <https://dspace.mit.edu/bitstream/handle/1721.1/47936/frameworkformana00gorr.pdf>

bieżący kontrolowanego przez siedzących w kabinie oficerów⁵⁹. Jednak przez długi czas było obecne tylko tam, gdzie decyzje do podjęcia dało się łatwo przełożyć na zestaw programowalnych zasad i warunków.

Przykładem dokonującego się postępu w zakresie zautomatyzowanego podejmowania decyzji może być branża nazywana obsługą klienta. Jej jakość jest podnoszona dzięki automatycznemu katalogowaniu rodzajów zgłaszanych zapytań, poprzez wprowadzenie *chatbotów* pomagających zebrać wszelkie dane potrzebne do obsługi zgłoszenia od klienta, aż po zaawansowane rozwiązania samodzielnie udzielające odpowiedzi⁶⁰.

Podobne technologie mogą znacząco poprawić naszą relację z państwem. Dzięki wykorzystaniu przez administrację automatyzacji, w tym automatyzację decyzji, obywatel jest w stanie szybciej uzyskać odpowiedź na zapytanie do urzędu lub dostać formularz wstępnie wypełniony przez automat. Wdrażanie ADM w administracji publicznej jest więc kolejnym krokiem we wdrażaniu wizji e-administracji: instytucji publicznych usprawnionych przez ich cyfryzację. Poza odpowiadaniem na zapytania obywateli technologia może w przyszłości stanowić rodzaj roboasystenta urzędnika, pomagając mu na przykład odnaleźć w przepastnych bazach danych państwa niezbędne do wypełnienia formularza informacje, zanim zwróci się on z prośbą o uzupełnienie do samego obywatela⁶¹.

Innym przykładem jest (częściowa) automatyzacja pracy biurowej. Robotyczna automatyzacja procesów (*Robotic Process Automation*, RPA) to przykład branży, w której oprogramowanie odciąża człowieka z najzwyklejszych i najbardziej powtarzalnych zadań. Ta współpraca ma potencjał stać się prawdziwą symbiozą. Robot może zająć się zadaniami powtarzalnymi, które można łatwo i precyzyjnie opisać w postaci procesu, jednocześnie unikając błędów popełnianych przez pracowników przy takich żmudnych zadaniach. Z kolei pracownik lub urzędnik będzie miał więcej czasu, który może poświęcić trudniejszym przypadkom – niecodziennym pytaniom czy skomplikowanym sprawom.

Coroczna ankieta Deloitte przeprowadzona wśród firm stosujących RPA⁶² pokazuje, że aż 86% menedżerów wskazało, iż roboty spełniły ich oczekiwania pod względem podnoszenia produktywności. Firmy, które dopiero zaczynają stosować automatyzację procesów, przewidują, że około 20% pracy uda się zautomatyzować. Z kolei firmy, które już stosują RPA, liczą na automatyzację połowy ich zadań.

⁵⁹ *How much does the autopilot actually fly*, <https://scandinaviantraveler.com/en/aviation/how-much-does-the-autopilot-actually-fly>

⁶⁰ *Iloraz sztucznej inteligencji - potencjał AI w polskiej gospodarce (edycja 2)*, <https://www.digitalpoland.org/assets/publications/iloraz-sztucznej-inteligencji/iloraz-sztucznej-inteligencji-edycja-2-2019.pdf>

⁶¹ *Algorithmic Government: Automating Public Services and Supporting Civil Servants in using Data Science Technologies*, <https://academic.oup.com/jnl/article/62/3/448/5070384>

⁶² *Robotic process automation (RPA) trends and to-do list for scaling across the enterprise*, <https://www2.deloitte.com/us/en/pages/operations/articles/global-robotic-process-automation-report.html>

Te liczby są zaledwie przewidywaniami dotyczącym wąskiej części gospodarki i jedynie automatyzacji procesów. Wpływ systemów ADM na całą gospodarkę i administrację będzie szerszy. Może to przynieść korzyść również społeczeństwu – obywatele mogą zyskać dostęp do usług państwowych i prywatnych o dowolnej godzinie dowolnego dnia, uzyskiwać odpowiedź praktycznie natychmiastowo i w przyszłości – liczyć na to, że państwo będzie wymagać od nas mniejszej ilości dokumentów. Zyski mogą być więc znaczące, nie tylko dla firm i gospodarki, lecz także – dla społeczeństwa.

Podjęmowanie decyzji w oparciu o dane i wiedzę

Kreowanie i wdrażanie strategii, tworzenie polityk publicznych czy codzienne podejmowanie decyzji to działania polegające w dużej mierze na selekcji istotnych czynników bądź argumentów spośród nieprzebranej ilości dostępnych informacji. Paradoksalnie, możemy wiedzieć za mało i za dużo równocześnie. Często brak nam informacji w kwestiach, które uznajemy za kluczowe, tonimy za to w informacjach pobocznych. Jeszcze więcej jest wiedzy – w postaci dostępnych w sieci danych, artykułów, analiz, badań i raportów, które niekoniecznie pozwalają nam skupić się na najważniejszych czynnikach. Na przykład, aby pozostać na bieżąco z publikowanymi artykułami dotyczącymi osiągnięć w medycynie, lekarz powinien dziennie czytać pięć tysięcy dokumentów⁶³.

Aby z niemożliwego do objęcia umysłem morza informacji pozyskać wiedzę, możemy wykorzystać algorytmy. Są one w końcu niestrudzone w analizowaniu informacji, co czynią zdecydowanie szybciej niż człowiek. Nie mają ograniczeń w liczbie równocześnie analizowanych dokumentów. Coraz lepiej radzą sobie z wyborem informacji dla nas kluczowej. Potrafią równocześnie przeanalizować tysiące lub miliony scenariuszy różniących się jedynie początkową decyzją i przedstawiać nam wyniki symulacji.

Zautomatyzowane podejmowanie decyzji odgrywa kluczową rolę w umożliwieniu wykorzystania gromadzonych przez nas danych. Dzięki pomocy systemów selekcyjnych informacje, przedstawiających podsumowanie czy symulujących scenariusze więcej decyzji może być podejmowane w oparciu o rzetelną, dogłębną analizę. Nie musimy opierać się wyłącznie na ludzkiej pamięci i doświadczeniu, które bywa zwodnicze, możemy wyposażać podejmujących decyzje i ekspertów w jeszcze szerszy przegląd wiedzy czy wyniki konkretnych symulacji. Co więcej, ich dostarczenie staje się tańsze i mniej czasochłonne.

Co istotne, tworzona w procesie wiedza może być aktualizowana na bieżąco z wykorzystaniem danych generowanych przez cały czas. Na przykład aplikacja zbierająca na żywo informacje z miejskich czujników ruchu i przewidująca miejsca przyszłych korków pozwala na wyprzedzającą reakcję. Nie tylko coraz więcej decyzji możemy

⁶³ Proliferations of Scientific Medical Journals: A Burden or A Blessing, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3191655/>

podejmować w oparciu o dane historyczne, ale też możemy analitykę stosować do gromadzonych na bieżąco informacji.

Obserwacja i eliminacja części ludzkich błędów i stronniczości

Coraz częściej w debacie publicznej wspomina się o stronniczości algorytmów (ang. *bias*), którą omówimy również poniżej. Jednak zebranie danych i wykorzystanie analizy do wnioskowania pozwala także dostrzec stronniczość i błędy człowieka. Czasami ludzkie uprzedzenia są znane wcześniej, czasem dopiero porównując wyniki działań ludzi do algorytmów, jesteśmy w stanie zaobserwować różnicę. Niekończenie jednak to działania algorytmu okazują się w takich sytuacjach błędne.

Jednym z przykładów jest analiza wyników oceny naukowców w Wielkiej Brytanii. Brytyjski system – w odróżnieniu od na przykład polskiego – jest w całości oparty o ocenę przez ekspertów czteroletniego dorobku publikacyjnego. W wielu innych krajach do takiej oceny używa się współczynnika cytowań utworów, który jest wskazówką ich wpływu na naukową debatę. Wykorzystanie cytowań spotyka się, również w Polsce, z żywą i niebezpieczną krytyką. Tymczasem według badań Alana Dixona⁶⁴, który porównał ocenę ekspercką do parametrycznej, brytyjscy eksperci wystawiali proporcjonalnie niższe oceny pracom wdrożeniowym w informatyce, a wyższe – teoretycznym. Przyczyn tego faktu może być wiele – jednym z proponowanych wytłumaczeń jest stereotypowe uznanie prac teoretycznych za trudniejsze i bardziej abstrakcyjne. Ten przypadek pokazuje, jak prosty algorytm, mimo swoich wad i sprowadzania dorobku naukowego do liczby cytowań, może okazać się mniej stronniczy od ekspertów.

Algorytmy dają nam okazję do obserwacji i zmierzenia się z ludzką stronniczością⁶⁵. Wspomniane powyżej wykorzystanie zautomatyzowanego podejmowania decyzji do odciążenia pracowników w powtarzalnych, żmudnych zadaniach to szansa na podniesienie produktywności, ale też zmniejszenia liczby błędów⁶⁶. Tworzenie systemów zautomatyzowanego podejmowania decyzji może więc doprowadzić nas do eliminacji części decyzji nieuprawnionych, dyskryminujących czy błędnych.

Wyzwania

Współpraca człowiek–algorytm

Automatyzacja podejmowania decyzji oznacza konieczność współpracy człowieka i komputera w jednym procesie. To jednak jest wyzwaniem, ze względu na różną

⁶⁴ *Citations and Sub-Area Bias in the UK Research Assessment Process*, <https://alandix.com/docs/ref2014/bib-2015-citations-and-ref-v2.pdf>

⁶⁵ *AI isn't dangerous, but human bias is*, <https://www.weforum.org/agenda/2019/01/ai-isn-t-dangerous-but-human-bias-is/>

⁶⁶ *Can Artificial Intelligence Help Reduce Human Medical Errors? – Economic and Policy Implications of AI*, <https://techpolicyinstitute.org/2018/07/17/can-artificial-intelligence-help-reduce-human-medical-errors-economic-and-policy-implications-of-ai-blog-post-4/>

naturę i charakterystykę procesów podejmowania decyzji przez ludzi i algorytmy. Algorytm nie działa jak „inny człowiek”, dlatego trudno nam zwrócić uwagę na jego potencjalne błędy. Współpraca człowieka z algorytmem może być wzajemnym uzupełnieniem, harmonijną symbiozą, ale charakterystyka algorytmów potrafi też być dla człowieka zwodnicza.

Człowiek w czasie swojej nauki naturalnie tworzy heurystyki, zasady bardziej instynktownego rozumowania pozwalającego radzić sobie z powtarzalnymi sytuacjami. W ten sposób możemy szybciej reagować na powtarzalne sygnały lub działać w powtarzalnych procesach, nie musząc angażować pełni naszych zdolności analitycznych, aby uczyć się reakcji od nowa za każdym razem. To znaczące ułatwienie, które jednak może nas zwieść na manowce.

W miarę używania wspomagających nas systemów komputerowych nasz podstawowy instynkt zaczyna sprowadzać się do hasła „zaufać maszynie”. Mamy szybko rosnące poczucie, że obecność algorytmu wprowadza w proces racjonalność. Łatwo damy się przekonać do niezawodności systemu zautomatyzowanego podejmowania (lub sugerowania) decyzji.

Zaufanie rośnie wraz z upływem czasu i liczbą przypadków, w których to maszyna przyczyniła się do właściwego rozwiązania. W końcu jako ludzie chcemy minimalizować wysiłek – na początku być może będziemy niezależnie sprawdzać, czy nasze własne rozumowanie doprowadzi nas do podobnych wniosków, ale z czasem zrezygnujemy z rygoru, widząc, że wnioski algorytmu zawsze pokrywają się z naszymi⁶⁷.

Wysokie zaufanie do systemu sugerującego decyzje powoduje, że w sytuacji konieczności podjęcia decyzji zazwyczaj staje się on naszym wiążącym kierunkowskazem. To zachowanie jest wzmacniane stronniczością człowieka, polegającą na ignorowaniu informacji niezgodnych z zaleceniem algorytmu lub postrzegania ich jako zgodnych (ang. *assimilation bias*) czy też poświęcanie zbyt dużej uwagi informacjom spójnym z tezą algorytmu (ang. *confirmation bias*)⁶⁸.

Problemem jest fakt, że systemy ADM są wysoko wyspecjalizowane, ale o wiele gorzej niż człowiek radzą sobie ze zmiennością otaczającego je świata. Są często lepsze niż my w analizowaniu przypadków, do których zostały przeszkolone. Nawet jednak niewielkie (z perspektywy człowieka) zmiany warunków zewnętrznych mogą zafałszować ich wnioskowanie.

Człowiek, posiadający inteligencję ogólną i elastyczną, niejako zakłada, że dla algorytmu również nie powinno to być problemem. Mając wspomniane wyżej zaufanie

⁶⁷ *Social Cognition*, https://books.google.pl/books/about/Social_Cognition.html?id=6Uq3QgAACAAJ&redir_esc=y

⁶⁸ *A Hypothesis-Confirming Bias in Labeling Effects*, <http://bear.warrington.ufl.edu/brenner/mar7588/Papers/darley-gross-jpsp1983.pdf>

⁶⁹ *Automation Bias in Intelligent Time Critical Decision Support Systems*, <https://web.archive.org/web/2014110113133/http://web.mit.edu/aeroastro/labs/halab/papers/CummingsAIAAbias.pdf>

do jego działania, sugeruje się więc jego rekomendacjami lub nadal pozwala mu decydować. To zjawisko nazywane jest stroniczością ze względu na automatyzację (*automation bias*)⁶⁹. Przekonany o racjonalności i nieomyślności maszyny człowiek potrafi ufać jej bardziej niż swoim umiejętnościom.

To zjawisko obserwowane było na przykład w przypadku autopilotów dwie dekady temu. Już w 1994 roku wskazywano, że ze 166 badanych incydentów lotniczych, w których rolę odgrywał autopilot, cztery na pięć zagrożeń lotu powstało w momencie spokojnego lotu na stałej wysokości, a w ponad połowie – autopilot działał dokładnie tak, jak został zaprogramowany. Zamiast jednak pomagać podnieść bezpieczeństwo lotu, doprowadzał do niebezpiecznych sytuacji, których nie wychwycili mający zbytne zaufanie do działania algorytmów piloci⁷⁰.

Kolejne badania pokazały, że ten efekt – nazywany utratą czujności z powodu automatyzacji – występuje w sytuacjach, w których człowiek wykonuje równocześnie zadania manualne i kontroluje te zautomatyzowane⁷¹. Musząc oszczędnie gospodarować swoją uwagą, pozostawia odpowiedzialność za zautomatyzowane procesy algorytmom. W testach systemu ostrzegania załogi samolotu o awariach silnika okazało się, że piloci wykrywali mniej awarii, pomagając sobie algorytmami, niż gdy musieli sytuację monitorować wyłącznie swoimi oczami. Z drugiej strony, obecność algorytmów pozwala oczywiście odciążyć pilota w natłoku równoległych zadań.

Jeszcze innym efektem wprowadzenia systemów zautomatyzowanego podejmowania i sugerowania decyzji jest sprawienie, że nowicjusze mogą pełnić rolę ekspertów. Zadania wymagające lat treningu, wypracowania wyczucia i zgromadzenia doświadczenia nagle mogą być – przy współudziale algorytmów – obsługiwane przez świeżo przeszkolonych ludzi. To oczywiście ma szereg pozytywnych konsekwencji: obniżenie bariery wiedzy pozwala większej liczbie osób na zdobycie bardziej zaawansowanych zawodów. Z drugiej strony, w sytuacjach krytycznych, wyjątkowych okolicznościach, brak wiedzy lub doświadczenia może doprowadzić do sytuacji niebezpiecznych lub szkodliwych, szczególnie gdy wziąć pod uwagę brak wszechstronności systemów ADM.

Sprawiedliwość

Systemy ADM mogą przyczynić się (i przyczyniają) do budowania społecznej niesprawiedliwości i pogłębiania nierówności poprzez powielanie stereotypów i generowanie szkodliwych ocen. Podstawowym źródłem takich skrzywień są wadliwe zbiory danych, na których trenuje się algorytmy. Wadliwe przede wszystkim z powodu nieobiektywnych ludzkich ocen, odzwierciedlonych na przykład w stosowanych takso-

⁷⁰ *Automation Use and Automation Bias*, <https://journals.sagepub.com/doi/abs/10.1177/154193129904300346>

⁷¹ *Complacency and Bias in Human Use of Automation: An Attentional Integration*, <https://journals.sagepub.com/doi/10.1177/0018720810376055>

⁷² *Deep Learning And The Limits Of Learning By Correlation Rather Than Causation*, <https://www.forbes.com/sites/kalevleetaru/2019/04/18/deep-learning-and-the-limits-of-learning-by-correlation-rather-than-causation/#3a8a141f7487>

nomiach danych. Ucząc się, systemy korzystają z danych historycznych zawierających nieobiektywne ludzkie oceny. Nawet jednak korzystając z odpowiednio dobranych danych, źle zaprojektowany algorytm może generować niesprawiedliwość⁷².

Pierwszym problemem związanym z potencjalną niesprawiedliwością są przypadki skrajne oraz „małe” kategorie, czyli takie, które obejmują niewiele przypadków. Może to dotyczyć niezwykle rzadkich chorób, mniejszości wyznaniowych czy innych niewielkich grup wyróżnionych pewną wspólną cechą, której ślad algorytm odnajdzie w danych. W takich sytuacjach algorytmy często przestają być skuteczne, co wynika po prostu z zasad statystyki.

Podstawowa ocena skuteczności algorytmu podczas jego treningu – jego precyzja – bierze pod uwagę dwa parametry: dokładność i wrażliwość. Dokładność (ang. *precision*) oznacza tutaj to, jak wiele podjętych przez algorytm decyzji jest poprawnych w odniesieniu do wszystkich podjętych przezeń decyzji. Innymi słowy odpowiada na pytanie: jak wiele z wybranych elementów (podjętych decyzji) jest poprawnych? Wrażliwość z kolei (inaczej odzyskanie, ang. *recall*) to liczba poprawnych rezultatów algorytmu podzielona przez całkowitą liczbę przypadków, w których algorytm powinien dać konkretny rezultat. Odpowiada więc na pytanie: jak wiele elementów wybrał algorytm w porównaniu do wszystkich elementów, które wybrać powinien?

Weźmy za przykład chorobę, która występuje u jednego na sto tysięcy pacjentów. Podczas tworzenia algorytmu diagnozującego choroby wpływ rzadkiej przypadłości na całkowitą ocenę dokładności i wrażliwości algorytmu będzie znikomy. Nawet jeśli u połowy pacjentów, którym dolega rzadkie schorzenie, nie postawi celnej diagnozy, będzie to stanowić zaledwie ułamek procenta wszystkich klasyfikowanych przez niego przypadków. Koniec końców możemy otrzymać więc algorytm, który z jednej strony celnie klasyfikuje ponad 99% pacjentów, a z drugiej w przypadku rzadkich chorób częściej się myli niż wyciąga celne wnioski.

Drugim aspektem, który może mieć wpływ na błędną ocenę przypadków rzadkich, jest niska różnorodność danych wejściowych. W kształceniu algorytmu występuje efekt zakotwiczenia i uprzedzenia (stronniczości) wynikły ze stabilizacji (*anchoring and stability bias*)⁷³. Oznacza to, że na samokształcący się algorytm największy wpływ mają dane początkowe, późniejsze zmiany mają bardziej charakter ewolucyjny niż rewolucyjny. Dlatego na przykład niereprezentatywna próbka danych lub brak dużej ilości danych może przełożyć się na jeszcze bardziej nieelastyczny lub błędnie rozumujący algorytm.

Wyzwaniem jest także homogeniczność osób przygotowujących dane i same systemy. Na przykład w przypadku technologii uczenia nadzorowanego dane są pre-klasyfikowane przez człowieka, co pozwala na rozpoczęcie kształcenia algorytmów. Zbyt duża homogeniczność (rasowa, kulturowa, etniczna, religijna) kategoryzujących

⁷³ Controlling machine-learning algorithms and their biases, <https://www.mckinsey.com/business-functions/risk/our-insights/controlling-machine-learning-algorithms-and-their-biases>

dane może poskutkować stworzeniem niereprezentatywnej nawet dla lokalnej społeczności klasyfikacji. Tymczasem w tym procesie, w zależności od algorytmu, kategoryzowanie dotyczy może nie tylko obiektywnych charakterystyk (jak na przykład rozpoznanie samochodu na zdjęciu), ale też subiektywnych odczuć (na przykład stwierdzenie czy człowiek na zdjęciu się uśmiecha lub wyraża radość).

Przykładem problemu może być choćby baza obrazów ImageNet służąca do kształcenia sztucznej inteligencji. W tym celu przechowywane obrazy zostały oznaczone przez człowieka, który opisywał (tagował) to, co znajduje się na obrazku. Dzięki analizie tysięcy zdjęć oznaczonych jako zawierające obraz kota, algorytm może sam „wymyślić”, jak wygląda kot – i samemu rozpoznawać go na kolejnych fotografiach. Okazało się jednak, że ludzie tagujący zdjęcia nie są wolni od swoich doświadczeń, stronniczości i stereotypów. Podczas eksperymentu nazwanego ImageNet Roulette ludzie zachęceni byli do wrzucenia do klasyfikatora swoich własnych zdjęć i pochwalenia się, jak zostali przez sztuczną inteligencję rozpoznani. Okazało się, że poza rozpoznaniem na zdjęciu „osoby”, „kobiety” czy „mediatorki”, ciemnoskórzy mężczyźni bywali rozpoznawani jako „przestępcy”, a dziewczyna z azjatyckimi rysami twarzy jako „jihadystka”⁷⁴. Co więcej, znane nam cechy dyskryminacyjne (jak płeć czy rasa) to jedynie czubek góry lodowej, który mamy przynajmniej dobrze rozpoznany. Co, jeśli któryś z rekrutacyjnych algorytmów uzna na przykład za sensowne zatrudnianie osób jedynie z angielskobrzmiącymi nazwiskami⁷⁵?

To jednak nie koniec wyzwań związanych ze sprawiedliwością algorytmu. Nawet największa ilość różnorodnych danych może zasugerować algorytmowi wykorzystanie czynników dyskryminujących, błędnych lub z nimi powiązanych. Może to dotyczyć cech, na które tradycyjnie zwracamy uwagę, jak płeć, kolor skóry, ale też zupełnie nowych opracowanych przez algorytm charakterystyk, jak choćby na przykład nazwisko kończące się na „-ski” czy polskobrzmiące imię. Zagrożeniem są więc szczególnie cechy pośrednie – w sposób nieoczywisty powiązane z cechami zazwyczaj kojarzonymi z dyskryminacją. Na przykład standardowo znajdująca się na CV informacja o miejscu zamieszkania może pozwalać na profilowanie według statusu materialnego.

Wreszcie nawet największy wysiłek na rzecz eliminacji stronniczości nie jest gwarantem sprawiedliwości działania, bo podejmowanie decyzji często zawiera element subiektywnej oceny, kierowanej wartościami podzielanymi przez osobę lub społeczność. Systemy ADM nie zawsze będą dysponowały obiektywnymi przesłankami do podjęcia lub sugerowania decyzji. Klasycznym dylematem są tutaj wybory samochodów autonomicznych, które dotyczą przedkładania ochrony pasażerów nad ochronę przechodniów lub osób młodszych nad starsze. To rozwinięcie klasycznego problemu – tak zwanego dylematu wagonika – stało się nawet tematem stworzonej

⁷⁴ *How ImageNet Roulette, an Art Project That Went Viral by Exposing Facial Recognition's Biases, Is Changing People's Minds About AI*, <https://news.artnet.com/art-world/imagenet-roulette-trevor-paglen-kate-crawford-1658305>

⁷⁵ *Hiring Algorithms Are Not Neutral*, <https://hbr.org/2016/12/hiring-algorithms-are-not-neutral>

⁷⁶ *Moral Machine*, <http://moralmachine.mit.edu/>

przez zespół MIT gry-ankiety nazwanej „Moralną Maszyną”⁷⁶. W niej samemu możemy się postawić przed problemami automatycznego kierowcy, który w mgnieniu oka ustalić ma wartość życia noworodka, osoby starszej, zwierzęcia domowego. Czy najechanie na wózek z noworodkiem jest lepszym, czy gorszym wyjściem od rozjechania emerytki? A dwóch emerytek?

Przejrzystość i wytłumaczalność

Dominująca jeszcze do niedawna narracja przedstawiała algorytmy jako czarne skrzynki: obywatelowi trudno jest zajrzeć w głąb systemów, które w niezrozumiały i często niewytłumaczalny dla nas sposób podejmują decyzje mające wpływ na nasze życie. Techniczne skomplikowanie powoduje, że nie tylko wytłumaczenie użytkownikowi działania kodu jest skomplikowane. Trudny i kosztowny staje się też audyt systemu.

Algorytm dodatkowo w czasie swojego działania potrafi wciąż ewoluować, wprowadzając potencjalnie konieczność przeprowadzania takiego przeglądu okresowo. W ciągu ostatnich lat oś debaty znacząco się jednak zmieniła, a naukowcy i eksperci intensywnie pracują nad otwieraniem czarnych skrzynek. Być może pewną drogą znalezienia rozwiązania jest wytłumaczalność samych wyników działania algorytmów⁷⁷ (pozwalająca na analizę rezultatów działania algorytmu) lub stworzenie procesu certyfikacji algorytmów na wzór certyfikacji nowych leków⁷⁸.

RODO było pierwszym znaczącym, ale wciąż niewystarczającym, krokiem prawnym służącym do zmiany tego stanu. Artykuł 22 RODO wprowadził zakaz podejmowania decyzji mających znaczący wpływ na człowieka w sposób całkowicie zautomatyzowany (tym samym wymagając jakiejś formy udziału człowieka w procesie), chyba że człowiek świadomie zaakceptuje pełną automatyzację. Z kolei artykuły 13–15 wprowadziły między innymi obowiązek wytłumaczenia logiki postępowania⁷⁹. Według obowiązującej interpretacji tych przepisów ubiegać się możemy za to o przedstawienie ogólnych kryteriów rozumowania stojących za podjętą wobec nas decyzją, bez ujawniania samego algorytmu⁸⁰. Nadal nie możemy jednak oczekiwać wytłumaczenia konkretnej decyzji podjętej w naszym przypadku.

Prowadzone są intensywne prace nad tak zwaną wytłumaczalną sztuczną inteligencją (ang. *explainable AI*)⁸¹. Jest to istotne szczególnie w przypadku systemów ADM opartych o metody głębokiego uczenia (*deep learning*). Metoda rozumowania takich systemów – przyjęte lub stworzone w modelu kategorie, cechy czy połączenia

⁷⁷ Detecting Bias: Does an Algorithm Have to Be Transparent in Order to Be Fair?, http://ceur-ws.org/Vol-2103/paper_1.pdf

⁷⁸ We should treat algorithms like prescription drugs, <https://qz.com/1540594/treating-algorithms-like-prescription-drugs-could-reduce-ai-bias/>

⁷⁹ Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling, <https://www.sciencedirect.com/science/article/pii/S026736491730376X>

⁸⁰ Article 29 Working Party, https://edpb.europa.eu/our-work-tools/article-29-working-party_en

⁸¹ Explanation in Human-AI Systems: A Literature Meta-Review, Synopsis of Key Ideas and Publications and Bibliography for Explainable AI, <https://arxiv.org/pdf/1902.01876.pdf>

między nimi – nie jest bowiem łatwo wyrażalna w zrozumiałych dla człowieka pojęciach i metodach rozumowania.

Globalizacja pomijająca lokalne wartości

Tak jak w innych sektorach internetowej gospodarki, musimy się liczyć z tym, że wąskie grono globalnych graczy zdominuje swoimi rozwiązaniami konkretne rynki systemów ADM. Będzie to rodzić wyzwania wynikające z braku konkurencji lub związane z nadzorem technologii. Działanie w tak wielkiej skali niesie też za sobą proporcjonalnie ogromną odpowiedzialność oraz trudności w dostosowaniu do lokalnej sytuacji.

To wynik działania tak zwanych efektów sieciowych. Kluczowym efektem jest tutaj wzrost wartości sieci w zależności od liczby jej użytkowników. Oryginalnie relacja ta – nazywana prawem Metcalfe’a – mówiła, że wartość sieci jest proporcjonalna do liczby jej członków podniesionej do kwadratu. Choć precyzyjniejsze określenie tej relacji jest różne w zależności od charakteru usług czy sieci⁸², sam efekt jest obserwowalny w wielu dziedzinach e-gospodarki.

W przypadku części systemów ADM większa liczba użytkowników pozwala zgromadzić większą ilość danych, wykształcić lepsze algorytmy czy szybciej niż konkurencja rozwijać produkt czy usługę. Widzimy już w wielu internetowych branżach, np. mediach społecznościowych czy e-handlu, że ten proces naturalnie prowadzi do wyłonienia się wąskiej grupy światowych liderów.

To z kolei stawia przed nami nowe wyzwania. Po pierwsze, nawet niewielka niezidentyfikowana stroniczość w produkcie lub usłudze będzie oddziaływać w ogromnej skali, potencjalnie na całym globie. Wyobraźmy sobie, że wspomniany wcześniej przykład błędu w algorytmie rekrutacyjnym firmy Amazon nie dotyczy jedynie zatrudnienia w jednej korporacji, ale w całej branży. Wówczas nawet niewielki błąd, który pozostawałby niewykryty, praktycznie mógłby eliminować całe grupy społeczne z rynku pracy poprzez znaczące ograniczanie ich możliwości zatrudnienia.

Dzisiaj ludzie dokonujący decyzji są w naturalny i indywidualny sposób stronniczy, jednak ze względu na różnice między nimi efekt niweluje się niejako samoistnie. Wyjątkiem są tutaj uprzedzenia dotyczące większości społeczeństwa, ale nawet wówczas możemy znaleźć część społeczności mającą inny światopogląd.

Generalizacja algorytmów – poszerzanie spektrum zadań, które algorytm może wykonywać czy też cech, które może brać pod uwagę – również jest potencjalnym zagrożeniem. Stosowana często w systemach ADM sztuczna inteligencja jest świetna w rozwiązywaniu konkretnych, wąsko określonych zadań, ale nadal zaskakująco (dla człowieka) ogromnym wyzwaniem jest realizacja usług o tylko minimalnie odmiennych założeniach początkowych. Nie bez powodu część ekspertów i badaczy wątpi w ogóle w możliwość stworzenia tak zwanej ogólnej (uniwersalnej) sztucznej inteli-

⁸² Metcalfe's Law is Wrong, <https://spectrum.ieee.org/computing/networks/metcalfes-law-is-wrong>

⁸³ The AI cargo cult: myth of superhuman AI, <https://www.wired.com/2017/04/the-myth-of-a-superhuman-ai/>

gencji, wielu również sądzi, że nawet w ostatnich latach nie udało nam się uczynić postępu w tym kierunku⁸³. Nieumiejętność radzenia sobie ze zmiennymi warunkami może powodować, że nawet świetnie wytrenowany algorytm będzie drastycznie tracił na skuteczności, a także wprowadzał stronniczość przy stosowaniu go w sytuacjach nawet bardzo podobnych, lecz różniących się szczegółami.

Różnica w społeczeństwach i kulturach prowadzi do różnic w światopoglądach, wadze argumentów, ale też moralności. Lokalność społeczeństw może niejednokrotnie stanąć w konflikcie z globalnością algorytmów narzucających jeden model całemu światu. Problem ten jest już widoczny w relacjach z technologicznymi gigantami internetu, gdy okazuje się, że zasady społeczności określone według standardów Doliny Krzemowej nie odpowiadają wrażliwości społeczeństwa na przykład w Polsce. Świetnym przykładem tego problemu są wyzwania z ustaleniem „standardów społeczności” sieci społecznościowych.

Prywatność

Trwająca debata o prywatności naszych danych jest blisko związana z kształconymi na nich systemami ADM. Kwestia prywatności jest kluczowym punktem wyjścia, koniecznym do dalszych prac nad zautomatyzowanym podejmowaniem decyzji.

Uwolnienie większej ilości danych mogłoby pozwolić na lepsze kształcenie algorytmów podejmujących decyzje. Większa ilość różnorodnych danych pozwoliłaby tworzyć modele biorące pod uwagę większą liczbę czynników. Sama skala dostępnych danych pozwoliłaby powiększyć (i potencjalnie zróżnicować) zestawy danych, na których kształcone są algorytmy⁸⁴, tym samym pozwalając uniknąć części wskazanych wyżej problemów. Wprowadzenie konieczności upubliczniania rezultatów działań algorytmów, czy też włączenie takich danych w proces certyfikacji pozwoliłoby zgromadzić rosnący zbiór danych służących do porównań różnych algorytmów.

Upublicznienie – bądź udostępnianie na określonych warunkach – większej liczby zbiorów danych niesie też za sobą możliwość deanonimizacji. Takie ryzyko rośnie też wraz z możliwością łączenia ze sobą różnorodnych zbiorów danych. Usuając dane umożliwiające naszą identyfikację ze zbioru danych, nie sprawiamy, że pozostałe informacje zostaną anonimowe na zawsze⁸⁵. Wykorzystując inne zbiory danych i łącząc zawarte w nich informacje, możemy dotrzeć do tożsamości konkretnej osoby. Na przykład dwóch naukowców z Uniwersytetu Texasu zdeanonimizowało dane udostępnione przez Netflix – pomimo usunięcia z nich identyfikatorów kont czy danych personalnych. Wykorzystali do tego IMDB, portal służący ocenie filmów. Łącząc te dwa zbiory, byli w stanie przypisać kontom z serwisu IMDB dokładną historię oglądania przez dane osoby filmów i seriali pobranych z Netflix⁸⁶.

⁸⁴ *Data Trusts*, <http://theodi.org/wp-content/uploads/2019/04/ODI-Data-Trusts-A4-Report-web-version.pdf>

⁸⁵ *Anonymous Data v. Personal Data — A False Debate: An EU Perspective on Anonymization, Pseudonymization and Personal Data*, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2927945

⁸⁶ *Robust De-anonymization of Large Sparse Datasets*, http://www.cs.utexas.edu/~shmat/shmat_oak08-netflix.pdf

Z kolei naukowcy z Uniwersytetu Louvain i MIT zbadali zanonimizowane dane dotyczące półtora miliona użytkowników pochodzące od operatora telefonii komórkowej z pewnego niewielkiego europejskiego państwa⁸⁷. Wykazali, że aby zidentyfikować 95% użytkowników, potrzeba zaledwie czterech punktów danych – konkretnych lokalizacji i czasu użycia telefonu w ciągu 12 miesięcy.

Anonimizacja nie jest więc jednorazowym działaniem, a ciągłym procesem. Każdy nowo pojawiający się zbiór danych niesie zagrożenie naszej deanonimizacji. Co więcej, zagrożenie takie będzie większe dla jednostek mających cechy charakterystyczne, takie jak rzadka przypadłość bądź choroba lub też miejsce zamieszkania poza wielką metropolią. To z kolei może przełożyć się na społeczne efekty, takie jak potencjalnie umożliwienie pożyczkodawcom i ubezpieczycielom dotarcia do znacznie większej ilości informacji na temat osób zamieszkujących na wsiach i w małych miejscowościach. Doprowadziłoby to do premiowania ich względem reszty mieszkańców kraju – na przykład poprzez niższe oprocentowanie kredytów – lub dyskredytacji – na przykład odmowy objęcia ubezpieczeniem.

Powiązany problem dotyczy wykorzystywania zasobów dostępnych publicznie – na przykład na wolnej licencji. W ich wypadku dylemat dotyczy nie tyle prywatności, co prawa do wykorzystywania danych osobowych jako surowca dla systemów ADM. W głośnym ostatnio przypadku IBM wykorzystał 700 000 wizerunków twarzy konkretnych osób, opublikowanych w serwisie Flickr, do stworzenia MegaFace, bazy danych służącej do uczenia systemów rozpoznawania twarzy. Zdjęcia te były dostępne na wolnej licencji umożliwiającej ich swobodne wykorzystywanie. Jednak przedstawione na zdjęciach osoby twierdzą, że nie wyraziły zgody na takie wykorzystanie ich wizerunków⁸⁸.

Kwestia prywatności jest nierozdzielnie związana z budową społecznej akceptacji dla systemów zautomatyzowanego podejmowania decyzji. Wypracowanie kontraktu społecznego zawierającego jasne reguły prywatności danych, czy też anonimizacji danych publicznie dostępnych stanowi punkt wyjścia do zdobycia społecznej akceptacji dla systemów ADM.

Odpowiedzialność

Pytanie o odpowiedzialność twórców bądź użytkowników systemu ADM jest fundamentalne w budowaniu społecznego zaufania do systemów zautomatyzowanego podejmowania decyzji. Najczęściej uwagę przykuwa pytanie o to, na kim spoczywa odpowiedzialność za decyzje autonomicznych samochodów – właścicieli pojazdu (który na kierunek jazdy nie ma wpływu) czy też jego wytwórcy? A dokładnie twórcy kodu zasilającego algorytmiczny system sterujący pojazdem.

⁸⁷ *Unique in the Crowd: The privacy bounds of human mobility*, <http://www.nature.com/articles/srep01376>

⁸⁸ *How photos of your kids are powering surveillance technology*, <http://www.nytimes.com/interactive/2019/10/11/technology/flickr-facial-recognition.html?action=click&module=Top%20Stories&pgtype=Homepage>

Przykład samochodu jest jednak jedynie jednym z wielu. Pytanie to musi powracać w każdym zastosowaniu zautomatyzowanego podejmowania decyzji. Każdy z systemów i każda z branż różnić się będzie zmiennością sytuacji – na przykład wpływem użytkownika na dane wejściowe lub oczekiwania co do wyniku, rodzajem stosowanych algorytmów. Problem szczególnie dotyczyć będzie algorytmów ciągle ewoluujących podczas ich użytkowania.

Już poruszone powyżej problemy ze sprawiedliwością algorytmów pokazują, jak skomplikowaną materią jest kwestia odpowiedzialności za poczynania zautomatyzowanego systemu. Z jednej strony jego działanie zależy bowiem od jakości danych wykorzystanych podczas kształcenia, ekstensywności testów i prób, dostępności danych porównawczych. Wszystkie te elementy sugerowałyby raczej konieczność wzięcia odpowiedzialności przez wytwórcę algorytmu, który często jako jedyny ma wgląd w proces tworzenia i kontrolę nad nim, również ze względu na tajemnicę firmy. Ten nie ma jednak pełnego przeglądu sytuacji, korzysta bowiem najczęściej z rozwiązań innych firm bądź społeczności *open source*, w postaci modułów, fragmentów oprogramowania, zewnętrznych serwisów i algorytmów dostępnych w formie usługi. Łączenie elementów w nowym systemie może skutkować wadami, których nie przewidział ani wytwórca „podzespołów” (który nie mógł przewidzieć każdego zastosowania swojego wytworu) ani twórca systemu ADM (który nie jest w stanie zweryfikować działania każdego elementu od samych podstaw).

Z drugiej strony efekty działania systemów ADM zależeć też będą od aktywności użytkownika, rozumianego tutaj jako osoba wykorzystująca algorytm do pomocy w podjęciu decyzji. Kluczowe będą stworzone przez niego lub nią polecenia, precyzja określenia oczekiwanego efektu działania, dostarczanie odpowiednio różnorodnych danych początkowych. Dlatego użytkownik – w tym my, korzystający z systemów ADM – również ponosi odpowiedzialność za efekt. Użytkownik nie jest w stanie zweryfikować od podstaw zasad działania systemu, ze względu choćby na brak czasu na przeglądanie tysięcy linii kodu czy brak dostępu do zasad jego działania objętych tajemnicą przedsiębiorstwa-wytwórcy. Dlatego trudno mu przewidzieć, które z zastosowań algorytmu jest już niepoprawne i może przynieść szkodliwy efekt.

Wreszcie trzecim aspektem jest funkcjonowanie systemów ADM w ekosystemie pełnym innych takich systemów. Coraz większa ich obecność w naszym życiu i gospodarce będzie nierozzerwalnie skutkować coraz częstszą interakcją lub wzajemnym wpływem dwóch (lub więcej) systemów zautomatyzowanego podejmowania decyzji. A to rodzi niebezpieczeństwo powstania niebezpiecznych lub nielegalnych interakcji i sprzężeń zwrotnych. Na przykład obserwowane już są zmywy cenowe algorytmów automatycznie ustalających ceny produktów na platformach e-handlu⁸⁹. Kto jednak powinien wziąć odpowiedzialność za takie przypadki, jeśli ich użytkownicy wcale nie nakazywali algorytmom podejmować działań nielegalnych, a jedynie działać w interesie firmy?

⁸⁹ *Left to Their Own Devices, Pricing Algorithms Resort to Collusion*, <http://www.popularmechanics.com/technology/robots/a26309827/left-to-their-own-devices-pricing-algorithms-resort-to-collusion/>

Wreszcie kolejnym elementem jest kwestia odpowiedzialności w przypadku weryfikacji procesu algorytmicznego podejmowania decyzji przez człowieka. Wiedząc o problemach ze współpracą na linii człowiek-komputer, które opisaliśmy powyżej, czy możemy obarczać pracownika odpowiedzialnością za każdą decyzję algorytmu? Takie podejście może skutkować znaczącym wzrostem ryzyka pracy z algorytmami, co z kolei opóźni ich wdrożenie. Brak wytłumaczalności przyczyn podjętej przez algorytm decyzji oraz jego zdolność do przetwarzania ogromnych ilości danych w ułamkach sekund powodują, że zdolność człowieka do weryfikacji poczynań systemów ADM i tak jest ograniczona. Czy sprawiedliwe jest więc czynienie ludzi odpowiedzialnymi za błędy podejmujących decyzje w ułamku sekundy, ewoluujących maszyn zbudowanych z milionów linii kodu?

Rekomendacje

Rozsądne i zrównoważone, określone jasnymi zasadami dbałości o interes publiczny wprowadzanie systemów zautomatyzowanego podejmowania decyzji może przynieść ogromne korzyści polskiej gospodarce, społeczeństwu i administracji publicznej.

Uniknięcie błędów i problemów popełnianych dzisiaj w innych państwach pomoże zmniejszyć społeczne, a także finansowe, koszty eksperymentalnych wdrożeń.

Jednocześnie dbałość o interes publiczny i budowanie zrozumienia wokół tematu ADM pomogą zbudować społeczne zaufanie do technologii, potrzebne, aby obywatele chcieli korzystać z innowacyjnych rozwiązań i wdrożeń.

Poniżej prezentujemy kluczowe zagadnienia, o które należy zadbać, chcąc wdrożyć w Polsce systemy ADM w taki sposób, by służyły one jednostkom i były zgodne z interesem publicznym. Przedstawiamy rekomendacje dotyczące pięciu kluczowych obszarów:

- 1) współpracy międzysektorowej,
- 2) zbiorów danych,
- 3) audytu,
- 4) regulacji,
- 5) edukacji twórców i użytkowników ADM.

W większości z tych obszarów nie ma prostych, gotowych rozwiązań – na świecie są natomiast inicjowane debaty nad ich pożądanym kształtem. Mamy nadzieję, że podobne dyskusje będą miały miejsce także w Polsce – wstępem do systemowej dyskusji o ADM była debata publiczna nad „Polityką rozwoju sztucznej inteligencji w Polsce na lata 2019–2027”⁹⁰.

Systemy ADM są tylko jednym elementem złożonego systemu społeczno-technologicznego, który powstaje na naszych oczach, w miarę jak wdrażamy różne technologie cyfrowe. Stanowią one kluczowy element tego systemu z perspektywy zabezpieczenia interesów społeczeństwa. Poszukiwanie pozytywnych rozwiązań w tym zakresie musi iść w parze z debatą nad pokrewnymi obszarami, takimi jak regulacja platform internetowych, decentralizacja infrastruktury sieciowej czy suwerenność cyfrowa. Przykładowo, przejrzystość algorytmów jest ważnym tematem w debatach

⁹⁰ Polityka Rozwoju Sztucznej Inteligencji w Polsce na lata 2019 – 2027 (projekt dla konsultacji społecznych), <https://www.gov.pl/attachment/0aa51cd5-b934-4bcb-8660-bfecb20ea2a9>

o regulacji platform internetowych i będzie jednym z głównych zagadnień dla obecnej Komisji Europejskiej⁹¹.

Powolny rozwój systemów ADM w Polsce czy zapóźnienia względem wiodących pod względem liczby wdrożeń państw (takich jak Stany Zjednoczone, Wielka Brytania czy kraje skandynawskie) mogą w tym przypadku stanowić dla Polski szansę. Mamy bowiem możliwość po raz kolejny wykorzystać rentę zapóźnienia – ucząc się na sukcesach i błędach innych, zadbać o odpowiednie wprowadzenie i regulację systemów ADM.

Jednocześnie musimy jako społeczeństwo przyjąć, że stosowanie systemów ADM jest tematem istotnym. Po pierwsze dlatego, że europejska debata o regulacji systemów ADM (a także szerzej – sztucznej inteligencji) już się toczy. Nie biorąc w niej udziału i nie przedstawiając własnego punktu widzenia, staniemy się przedmiotem polityki technologicznej, a nie podmiotowym jej twórcą.

Współtworząc i wdrażając rozwiązania, możemy zminimalizować straty i zmaksymalizować zyski ze stosowania systemów zautomatyzowanego podejmowania decyzji.

Nawet jeśli w Polsce niewiele jest wdrożeń takich systemów, to na naszych obywateli już dziś oddziałują algorytmy wykorzystywane przez zagraniczne podmioty, np. te stosowane w sektorze finansowym lub na platformach społecznościowych. Odpowiednia regulacja działania tych algorytmów powinna więc być przedmiotem naszej troski.

Pisząc o rozwiązaniach, mamy na myśli zarówno legislację, jak i inne środki regulacyjne kształtujące wykorzystanie algorytmów: standardy, kodeksy dobrych praktyk, samoregulację przez podmioty tworzące rozwiązania ADM. Nie możemy jednak zapominać, że niezwykle ważne są rozwiązania pozaregulacyjne, takie jak wypracowywanie nowych modeli biznesowych czy edukacja – zarówno użytkowników, jak i administracji publicznej.

Obszary i propozycje rozwiązań

Realna debata i współpraca – model poczwórnej helisy

Współpraca w modelu poczwórnej helisy oznacza współpracę wszystkich sektorów przy wdrażaniu i kształtowaniu systemów ADM. Uczestnictwo strony społecznej – której nie uwzględniają tradycyjne modele innowacji międzysektorowej – jest niezbędne, jeśli chcemy tworzyć polityki mające na celu nie tylko rozwój gospodarczy czy optymalizację działań administracji publicznej, lecz także dobrostan społeczny i prawa jednostek. Współpraca taka nazywana jest modelem poczwórnej helisy, co

⁹¹ A Union that strives for more, http://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_en.pdf

odnosi się do splątania czterech grup w sposób podobny temu, w jaki nasze DNA tworzy podwójną helisę⁹².

Maksymalizacja szans związanych z wykorzystaniem technologii, przy jednoczesnej redukcji towarzyszących im zagrożeń, wymaga współpracy i porozumienia pomiędzy wymienionymi grupami.

Działanie w modelu poczwórnej helisy jest niezbędne, jeśli chcemy zapewnić zrównoważoną podaż i popyt na technologie cyfrowe.

Chcemy, aby polska gospodarka się rozwijała i stawiała się coraz bardziej konkurencyjna, oraz aby sprawność działania naszego państwa się podnosiła. Jednocześnie interes społeczny powinien być równoważnym czynnikiem decydującym o kierunku rozwoju technologicznego, w tym wykorzystania systemów ADM. Ciągła weryfikacja skutków wprowadzania w życie innowacji wymaga nie tylko działań na poziomie oceny skutków prowadzonych przez podmioty wdrażające, lecz także badań naukowych i niezależnego spojrzenia organizacji społecznych. Takie podejście nie może ograniczać się jedynie do prowadzenia konsultacji społecznych, które niestety nadal często mają charakter symboliczny. Stanem idealnym jest ujęcie wszystkich wymienionych interesariuszy w procesie tworzenia, wdrażania czy weryfikacji rozwiązań technologicznych⁹³.

Skuteczne wdrożenie takiej kooperacji wymaga wypracowania odpowiednich modeli współpracy międzysektorowej. Przykładem może być holenderski Techniekpakt⁹⁴ – międzysektorowe porozumienie na rzecz zwiększenia podaży talentu cyfrowego. Potrzeba również instytucji zaprojektowanych tak, by wspierać współpracę w poczwórnej helisie. W tym przypadku wzorować można się na brytyjskim Open Data Institute⁹⁵ inicjującym współpracę na styku sektorów, dotyczącą nowych modeli zarządzania danymi. Międzysektorowy charakter powinny mieć też ciała powołane do wypracowania strategii rozwoju systemów ADM w społeczeństwie.

Model regulacji

Systemy ADM nie mogą zostać pozostawione bez regulacji – ze względu na wpływ, który mają na jednostki i całe społeczeństwo, w tym potencjalne zagrożenia. Pytanie jedynie, jaki jest niezbędny model i poziom regulacji. „Linią demarkacyjną” powinien być właśnie wpływ systemów ADM na jednostkę czy interes społeczny. W przypadku wywierania takiego wpływu niezbędna jest regulacja zapewniająca mechanizmy odwoławcze, a także kontrolę ze strony państwa (przy założeniu, że w społeczeństwie demokratycznym zadaniem państwa jest ochrona praw podstawowych).

⁹² „Mode 3’ and ‘Quadruple Helix’: toward a 21st century fractal innovation ecosystem, <https://www.inderscience.com/info/inarticle.php?artid=23374>

⁹³ Smart specialization in regional innovation systems: a quadruple helix perspective, <https://onlinelibrary.wiley.com/doi/pdf/10.1111/radm.12306>

⁹⁴ Techniekpakt, <http://techniekpakt.nl/>

⁹⁵ Open Data Institute, <http://theodi.org/>

Projektowanie systemów ADM musi opierać się nie tylko na założeniu redukcji szkód społecznych (na czym często skupia się regulacja), lecz także powinno umożliwiać wykorzystywanie szans na pozytywną zmianę społeczną. Debata o kompleksowym podejściu regulacyjnym do systemów ADM trwa od lat, jednak – jak piszą autorzy przeglądowego raportu *algo:aware. Raising awareness on algorithms* – wciąż istnieje niewiele konkretnych inicjatyw regulacyjnych. Autorzy jednocześnie przypominają, że stosowanie systemów ADM podlega istniejącym przepisom dotyczącym m.in. ochrony praw podstawowych, ochrony konsumentów czy norm bezpieczeństwa.

Projektując regulację systemów ADM, trzeba będzie więc zdecydować, w jakich sytuacjach konieczne są działania legislacyjne, a w jakich wystarczą jedynie mechanizmy *soft law* i pokrewne środki regulacyjne.

Twarda regulacja jest niezbędna wszędzie tam, gdzie istnieje ryzyko naruszenia praw podstawowych obywateli.

Mechanizmy samoregulacji przez podmioty rynkowe są w takich przypadkach niewystarczające. Krytyka samoregulacji dotyczy w szczególności granic oraz skuteczności tworzonych w ten sposób ram etycznych. Poza tym nadmierne poleganie na samoregulacji może doprowadzić do prywatyzacji prawa: przerzucania zadań takich jak egzekwowanie prawa na podmioty biznesowe. Firmy często nie zapewniają skutecznych ścieżek odwołania lub dochodzenia swoich praw – te muszą być zagwarantowane przez przepisy prawa.

Eksperti zajmujący się systemami ADM ciągle poszukują właściwego modelu regulacyjnego. Często proponowanym punktem odniesienia jest system certyfikacji leków – ze względu na porównywalną skalę zarówno pozytywnych skutków i potencjalnych zagrożeń. U progu XX wieku stworzono odpowiednie instytucje certyfikacji dla leków, takie jak amerykańska Federal Drug Administration⁹⁶, oraz wieloetapowy proces badań klinicznych. Jego celem jest z jednej strony udowodnienie przez wytwórcę, że substancja faktycznie działa i pozwala leczyć. Z drugiej strony celem wielu rund prób klinicznych jest zbadanie potencjalnych skutków ubocznych.

Porównanie leków do algorytmów nie jest jednak tak proste: większość systemów zautomatyzowanego podejmowania decyzji nie ma potencjalnie tak poważnych bezpośrednich skutków ubocznych, jak podanie pacjentowi niezbadanej substancji chemicznej. Ponadto zbadanie efektu działania leku jest procesem chemiczno-biologicznym, który jest łatwo weryfikowalny. Wykorzystanie algorytmów i jego skutki stanowią z kolei procesy społeczne zależne od wielu zmiennych, często trudnych do przewidzenia. Należy pamiętać, że proces certyfikacji leków jest długi i kosztowny –

⁹⁶ *The History of FDA's Fight for Consumer Protection and Public Health*, <https://www.fda.gov/about-fda/history-fdas-fight-consumer-protection-and-public-health>

potrafi trwać nawet dekadę, kosztować setki milionów dolarów⁹⁷, a taki model w sferze rozwoju technologii rodzi zagrożenia, takie jak nadużywania pozycji dominującej, eksploatacja danych, nierówność pozycji pomiędzy dostawcą algorytmu a użytkownikiem.

Co więcej, samouczące się algorytmy zmieniają swoją charakterystykę w czasie, przekształcając swój model poprzez naukę na błędach i nowych danych. Inaczej niż w przypadku leków – skutki uboczne mogą więc zmieniać się i ewoluować, nie da się ich określić na stałe przed wprowadzeniem algorytmu do użycia.

Jak widzimy, klasyczne modele regulacyjne nie zawsze sprawdzają się w przypadku rozwoju technologii, dlatego niezbędne jest poszukiwanie innych – często mieszanych – modeli regulacyjnych, co ma miejsce właśnie w przypadku systemów ADM. Część z proponowanych modeli opiera się na powołaniu nowej instytucji-regulatora (co wynika z założenia, że zarówno systemy ADM, jak i algorytmy wymagają specjalistycznej wiedzy i kompetencji), inne przypisują odpowiedzialność za regulację instytucjom już istniejącym (np. podmiotom odpowiedzialnym za ochronę danych osobowych).

Odmiernym podejściem jest próba regulacji poprzez stworzenie systemu (procesu) weryfikacji algorytmów, za którego wykonanie odpowiadać ma wytwórca lub wdrażający system ADM. Taki model wdrożony jest w Rozporządzeniu o Ochronie Danych Osobowych – podmiot wdrażający system musi spełniać określone warunki i gwarancje, a podmiot kontrolujący ma jedynie za zadanie sprawdzanie dopełnienia tych standardów.

Sam proces weryfikacji (czy też certyfikacji) systemu ADM również może przebiegać w wielu wariantach. Taka certyfikacja może odbywać się jeszcze przed wprowadzeniem rozwiązania do użycia na dużą skalę, czyli po fazie eksperymentalnej (co być może zamieniłoby proces trwający miesiące w taki trwający lata, tym samym zmniejszając do wdrażania innowacji). Może też być prowadzona okresowo, ale już po wdrożeniu systemu (co mogłoby narazić część społeczeństwa na niewykryte we wcześniejszych stadiach skutki uboczne).

Niezależnie od tego, jaki model regulacyjny zostanie przyjęty, kluczowe będzie zapewnienie transparentności, wyjaśnialności, a także możliwości odwołania od decyzji podjętej przez system ADM – co stanowi kolejny punkt naszych rekomendacji.

Wyjaśnienia, audyt i możliwość odwołania od decyzji

Przejrzystość działania jest fundamentem budowania zaufania do systemów ADM, szczególnie tych stosowanych przez administrację publiczną. Pozwala bowiem zrozumieć oddziaływanie tych systemów na jednostki, a więc też ogólny wpływ na spo-

⁹⁷ *Estimating the clinical cost of drug development for orphan versus non-orphan drugs*, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6327525/>

leczeństwo. Najwyższa pora uznać, że algorytmy nie są czarnymi skrzynkami (*black boxes*), podejmującymi decyzje, których nie można w żaden sposób i na żadnym poziomie wyjaśnić⁹⁸.

Metafora czarnej skrzynki sprawia, że godzimy się, by kluczowe systemy wpływające na społeczeństwo działały w sposób nieprzejrzysty, a przez to niepodlegający kontroli społecznej. To założenie niedopuszczalne, ale też coraz mniej prawdziwe. Powinniśmy oczekiwać wobec twórców oraz podmiotów wdrażających systemy ADM, by wyjaśniały sposób ich działania. Kompleksowa ocena skutków wdrożenia systemów ADM powinna być normą szczególnie w sektorze publicznym. Przykład modelu takiej oceny przedstawiła Fundacja AI NOW w raporcie *Algorithmic Impact Assessments*⁹⁹.

Systemy ADM powinny być przejrzyste – mogą być wyjaśniane poprzez odpowiednie procedury oceny skutków ich działania.

To wiąże się z kwestią odpowiedzialności za działanie technologii. Zagwarantowanie możliwości wglądu w system ADM pozwala choć częściowo wyrównać dysproporcję relacji władzy, którą generują systemy ADM – obserwujące nas, a jednocześnie często niemogące być przedmiotem obserwacji. Audyty i analizy dokonywane, a następnie udostępniane publicznie, przez podmioty wdrażające ADM zapewnią dużo większą przejrzystość niż sytuacja, w której dbają o nią jedynie eksperci, dziennikarze i działacze społeczni.

Wyjaśnialność algorytmów pozwala również skorygować wszelkie skrzywienia związane z ich funkcjonowaniem – skutkujące dyskryminacją i wzmacnianiem nierówności społecznych. Najwyższe standardy powinny przy tym oczywiście dotyczyć wdrożeń systemów ADM w sektorze publicznym.

Należy wypracować standardy wewnętrznej oceny skutków ADM. Jednocześnie obywatele, naukowcy, organizacje społeczne powinny również móc dokonać audytu stosowanych systemów ADM, stanowiąc kolejny element tworzonego systemu *checks and balances*. Podstawowym narzędziem do tego powinien być rejestr wdrożonych systemów ADM prowadzony przez organizacje wykorzystujące takie systemy. W ramach takiego rejestru instytucje wdrażające ADM określałyby cel wdrożenia, wykorzystywane dane wejściowe, rodzaj algorytmu, proces, w którym system jest wykorzystywany, potencjalne zagrożenia powstałe na skutek błędów w działaniu systemu oraz podjęte środki zaradcze.

Zobowiązanie do stworzenia, upublicznienia i okresowej aktualizacji takiego rejestru powinno objąć na pewno instytucje publiczne, a być może także spółki skarbu pań-

⁹⁸ *Explanation in Human-AI Systems: A Literature Meta-Review, Synopsis of Key Ideas and Publications and Bibliography for Explainable AI*, <https://arxiv.org/pdf/1902.01876.pdf>

⁹⁹ *Algorithmic Impact Assessments. A Practical Framework for Public Agency Accountability*, <https://ainowinstitute.org/aiareport2018.pdf>

stwa lub firmy prywatne, których działalność może mieć potencjalnie istotny wpływ na społeczeństwo bądź prawa jednostek (jak np. firmy ubezpieczeniowe czy banki).

Prawo powinno gwarantować odpowiednie mechanizmy odwoławcze i naprawy szkód.

Wyjaśnialność działania systemów ADM wiąże się też z kwestią możliwości odwołania się od algorytmicznej decyzji i zapewnienia wyrównania ewentualnych krzywd. Ścieżka odwoławcza może być zagwarantowana w różnych modelach, obejmujących nie tylko drogę sądową czy administracyjną, lecz także mediacje czy alternatywne metody rozwiązywania sporów. Zdaniem brytyjskiej organizacji Doteveryone.uk zapewnienie możliwości odwołania się powinno być jednym z zadań nowego podmiotu – Urzędu Odpowiedzialnych Technologii (Office for Responsible Technology)¹⁰⁰. Zapewnienie wyjaśnialności algorytmów wraz z możliwością odwołania znacząco zwiększyłoby kontrolę społeczną nad systemami ADM.

Zarządzanie danymi

Funkcjonowanie systemów ADM wymaga dostępu do odpowiednich danych – decyzje podejmowane przez systemy ADM będą jedynie tak dobre, jak jakość wykorzystywanych przez nie danych (co jest również tematem dyskusji o rozwoju sztucznej inteligencji). Odpowiednia jakość oraz przede wszystkim dostępność zbiorów danych jest też niezbędna, jeśli chcemy unikać skrzywionych i stronniczych wyników działania systemów ADM.

Dlatego też wdrożenie w Polsce systemów ADM wymaga pogłębionej dyskusji nad modelami zarządzania danymi. Punktem wyjścia powinien być model otwartych danych, wdrażany w Polsce od dekady – w jego ramach zdefiniowano potencjał rozwojowy związany z ponownym wykorzystaniem swobodnie dostępnych danych¹⁰¹. To model nadal nie w pełni wdrożony, bowiem wiele kluczowych zbiorów danych publicznych jest nadal niedostępnych.

Ważnym sygnałem w kwestii otwartości danych są rozwiązania proponowane w nowej „Polityce rozwoju Sztucznej Inteligencji w Polsce”. Opowiadamy się za postulowaną przez polski rząd zasadą, że nieprzetworzone dane powinny być otwarcie dostępne – dotyczy ona danych publicznych, pochodzących z otoczenia środowiska naturalnego człowieka oraz tych generowanych przez aktywność ludzką¹⁰². Otwarcie zbiorów danych nieosobowych jest też kluczem do społecznie odpowiedzialnego wdrażania systemów ADM, stanowi bowiem niezbędny warunek zapewnienia przejrzystości działania algorytmów.

¹⁰⁰ *Regulating for Responsible Technology: Capacity, Evidence and Redress: a new system for a fairer future*, <https://doteveryone.org.uk/project/regulating-for-responsible-technology/>

¹⁰¹ *Rynek produktów, usług i treści cyfrowych opartych na ponownym wykorzystaniu informacji sektora publicznego (ISP) w Polsce*, http://www.polskacyfrowa.gov.pl/media/1076/POPC_WISECC_ISP_raportkoncowy_2112015.pdf

¹⁰² *Polityka Rozwoju Sztucznej Inteligencji w Polsce na lata 2019 – 2027*

Należy wypracować modele zarządzania danymi pozwalające jednocześnie chronić prywatność, uwzględniać specyfikę danych osobowych i nieosobowych oraz umożliwić skuteczne wykorzystywanie danych przez różne podmioty (także komercyjne).

Dodatkowe warunki mogą zostać wprowadzone w przypadku danych generowanych lub zbieranych przez podmioty komercyjne. Jednak nawet w ich wypadku należy rozważyć różne modele otwartości danych. Pod hasłem *data trusts* powstają dzisiaj eksperymentalne struktury prawne umożliwiające współdzielenie danych w jednej bazie na zasadach określonych przez interesantów z sektora prywatnego i publicznego¹⁰³. Taka instytucja ma za zadanie zarządzanie danymi w konkretnej dziedzinie, na przykład geolokalizacyjnymi danymi miasta. Celem trustu jest pogodzenie interesów wytwórców danych, a także osób indywidualnych, instytucji oraz firm chcących z nich korzystać. W ten sposób umożliwia się wielokrotne wykorzystanie danych oraz tworzy jasne reguły korzystania z nich, jednocześnie zabezpieczając interes firm i instytucji tworzących zbiory danych. *Data trust* umożliwia pogodzenie interesów podmiotów posiadających dane oraz interesu publicznego, którym jest ich powtórne wykorzystanie w nowych zastosowaniach.

Najtrudniejsza do uregulowania pozostaje kwestia danych osobowych, szczególnie że są one dzisiaj zbierane i wykorzystywane przede wszystkim przez podmioty komercyjne. Rozpoczynające się prace nad europejskim Digital Services Act tworzą warunki do zapoczątkowania prac nad wypracowaniem odpowiednich, zrównoważonych modeli wykorzystywania danych osobowych, z poszanowaniem zasad stworzonych przez Rozporządzenie o Ochronie Danych Osobowych. Modele te powinny zakładać zagwarantowanie obywatelowi praw do swoich danych, tym samym zapewniając większą suwerenność jednostki w decydowaniu o ich wykorzystaniu. Nie chodzi tu przy tym jedynie o ochronę prywatności jednostki, ale też zachowanie społecznej kontroli nad sposobami wykorzystania danych osobowych, a więc działanie na rzecz interesu społecznego.

Edukacja

Żaden z poruszonych powyżej punktów nie przyniesie efektu, jeśli nie pogłębijmy wśród społeczeństwa zrozumienia tematu systemów zautomatyzowanego podejmowania decyzji oraz technologii, na których te systemy są oparte. W społeczeństwie kształtowanym coraz bardziej przez algorytmy podstawowa alfabetyzacja algorytmiczna jest koniecznością.

Edukacja – często wspominana, choć równie często bagatelizowana – to kluczowy element zrównoważonego rozwoju systemów ADM.

To od zrozumienia przez obywateli szans, wyzwań i sposobu działania systemów zautomatyzowanego podejmowania decyzji zależy, czy Polska zadba o popyt na rozwią-

¹⁰³ Defining a 'data trust', <https://theodi.org/article/defining-a-data-trust/>

zania ADM i doprowadzi do szybszego oraz zrównoważonego rozwoju technologii w naszym kraju.

Jednocześnie w obszarze edukacji, jak w żadnym innym, nie możemy szukać nowego otwarcia – kompetencje potrzebne do rozumienia i pracy z systemami ADM, algorytmami i sztuczną inteligencją to w dużej mierze te same kompetencje cyfrowe, o których rozwój strona społeczna zabiega od dwudziestu lat.

Podnoszenie kompetencji algorytmicznych musi być wielopoziomowe. Z jednej strony konieczne jest, by dotyczyło pogłębiania poziomu podstawowych umiejętności cyfrowych wśród obywateli. Jak pokazuje raport firmy SpotData oparty o dane Eurostatu¹⁰⁴, Polska odstaje od średniej europejskiej, a nawet od kraju regionu (jak Czechy) w statystyce odsetka obywateli posiadających przynajmniej średnie kompetencje cyfrowe. Niedostateczny poziom tych umiejętności to hamulec nie tylko dla rozwoju polskiej gospodarki opartej o technologie¹⁰⁵, lecz także dla budowania społecznego zaufania i społecznej kontroli nad systemami ADM.

Wśród kadry zarządzającej średniego i wysokiego szczebla wymagania dotyczące rozumienia tematu systemów ADM powinny być znacząco wyższe. Wspomniane powyżej pomysły nie będą możliwe do wprowadzenia w życie, jeśli pracownicy firm prywatnych i urzędnicy państwowi, wdrażający i wykorzystujący te narzędzia, nie będą rozumieć szans i zagrożeń wynikających ze stosowania takich systemów. Bez ogólnej bazy wiedzy i umiejętności zastosowania jej w przypadku konkretnych systemów, z którymi będą mieć na co dzień do czynienia, niemożliwa stanie się ocena społecznego wpływu czy wewnętrzna ewaluacja ryzyka.

Wreszcie rozumienie związanej z systemami ADM strony etycznej i społecznej stosowania takich systemów konieczne jest wśród „talentu cyfrowego”: specjalistów IT oraz ekspertów. Ci pierwsi są bowiem odpowiedzialni za tworzenie większości rozwiązań algorytmicznych – ich świadomość skutków społecznych i etycznych może wpływać na ostateczny kształt technologii. Z kolei eksperci będą odpowiadać za dokumentację systemów ADM, ich weryfikację (np. w procesie audytu lub podczas ewentualnej certyfikacji), czy też będą tworzyć polityki publiczne, regulaminy czy zasady wdrożeń sektorowych (np. w sądownictwie czy służbie zdrowia).

Ostateczną kluczową grupą są liderzy zmiany technologicznej – to przede wszystkim przedsiębiorcy i innowatorzy tworzący rozwiązania ADM. To od ich wiedzy i świadomości zależy, jakie rozwiązania będą wdrażane, na przykład w polskim sektorze publicznym. Polscy programiści i polskie startupy tworzą niewątpliwie historię sukcesu rozwoju polskiej gospodarki. Mamy nadzieję, że będą firmami, które chcą działać w sposób etyczny i świadomy społecznie, mając na uwadze więcej niż tylko zysk ekonomiczny i konkurencyjność na rynku.

¹⁰⁴ Dobrzy informatycy to nie wszystko, sami nie zmieniają gospodarki, <https://spotdata.pl/blog/2019/05/10/dobrzy-informatycy-to-nie-wszystko-sami-nie-zmienia-gospodarki/>

¹⁰⁵ Sztuczna inteligencja nie zbawi polskiej gospodarki, <https://klubjagiellonski.pl/2019/07/05/sztuczna-inteligencja-nie-zbawi-polskiej-gospodarki/>

Kluczowa literatura

Poniżej przedstawiamy najważniejsze raporty i analizy systemów ADM opracowane w ostatnich latach.

Algo:aware (2018). *Algo:aware. Raising awareness on algorithms*, European Commission's Directorate-General for Communications Networks, Content and Technology, <https://platformobservatory.eu/app/uploads/2019/06/AlgoAware-State-of-the-Art-Report.pdf>

Engin, Z., Treleaven, P.C. (2019). *Algorithmic Government: Automating Public Services and Supporting Civil Servants in using Data Science Technologies*, Computer Journal, 62, 448-460, <https://academic.oup.com/comjnl/article/62/3/448/5070384>

Izdebski, K., Škop, M., Merényi, M., Turashvili, T., Kerekeš, D., Ilić, V. (2019), *alGO-Vrithms - State of play: Report on Algorithms Usage in Government-Citizens Relations in Czechia, Georgia, Hungary, Poland, Serbia and Slovakia*, Fundacja ePaństwo, <https://epf.org.pl/en/wp-content/uploads/sites/3/2019/05/alGOVrithms-State-of-Play-Report.pdf>

Matzat, L. et al. (2019) *Atlas of Automation - Automated decision-making and participation in Germany*, AlgorithmWatch, https://atlas.algorithmwatch.org/wp-content/uploads/2019/04/Atlas_of_Automation_by_AlgorithmWatch.pdf

Miller, C., Ohrvik-Stott, J., Coldicutt, R. (2018), *Regulating for Responsible Technology: Capacity, Evidence and Redress: a new system for a fairer future*, Doteveryone, <https://doteveryone.org.uk/project/regulating-for-responsible-technology/>

Mueller, S., Hoffman, R., Clancey, W., Emrey, A. (2019), *Explanation in Human-AI Systems: A Literature Meta-Review, Synopsis of Key Ideas and Publications and Bibliography for Explainable AI*, DARPA XAI Literature Review, <https://arxiv.org/pdf/1902.01876.pdf>

Reisman, D., Schultz, J., Crawford, K., Whittaker, M. (2018), *Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability*, AI Now Institute, <https://ainowinstitute.org/aiareport2018.pdf>

Spielkamp, M. et al. (2019), *Automating Society: Taking Stock of Automated Decision-Making in the EU*, AlgorithmWatch we współpracy z Fundacją Bertelsmanna, https://www.ivir.nl/publicaties/download/Automating_Society_Report_2019.pdf

Staniłko, J. F., Kroplewski, R. et al. (2019), *Polityka Rozwoju Sztucznej Inteligencji w Polsce na lata 2019–2027, projekt dla konsultacji społecznych*, <https://www.gov.pl/attachment/0aa51cd5-b934-4bcb-8660-bfecb20ea2a9>

West, S.M., Whittaker, M., Crawford, K. (2019), *Discriminating Systems: Gender, Race and Power in AI*, AI Now Institute, <https://ainowinstitute.org/discriminating-systems.html>

Whittaker, M. et al. (2018), *AI Now Report 2018*, AI Now Institute, https://ainowinstitute.org/AI_Now_2018_Report.pdf

Fundacja Centrum Cyfrowe

Naszym celem jest wspieranie otwartości i zaangażowania w świecie nowoczesnych technologii. Działamy w modelu think-and-do tank, wspierając otwartość i zaangażowanie w świecie nowoczesnych technologii. Pracujemy na rzecz zmiany społecznej i zwiększenia zaangażowania obywatelskiego, wykorzystując narzędzia cyfrowe i modele współpracy oparte na dzieleniu się zasobami i wiedzą.

Naszym zdaniem zdrowe społeczeństwo cyfrowe musi uwzględniać społeczny wymiar technologii – to założenie, które przyświeca nam od początku działalności. Nie możemy skupiać się jedynie na rozwijaniu infrastruktury, gospodarce cyfrowej lub rozwoju usług i produktów. Zamiast o technologiach chcemy rozmawiać o SpołTechu.

Strona www: www.centrumcyfrowe.pl

Twitter: @centrumcyfrowe

Nasze główne obszary działania to:

Kompetentni użytkownicy. Wspieramy użytkowników technologii cyfrowych w podnoszeniu ich umiejętności i kompetencji związanych z otwartością i współpracą.

Otwarte instytucje. Współpracujemy z instytucjami, aby upewnić się, że działają one w sposób otwarty, aby realizować swoją misję społeczną.

Dobre przepisy. Pracujemy nad dostosowaniem przepisów i wykorzystaniem narzędzi prawnych w celu wspierania potrzeb i praw użytkowników, którzy uczestniczą w otwartym obiegu zasobów online.

Społeczny wymiar technologii. Diagnozujemy zmiany społeczne i kulturowe zachodzące w naszym społeczeństwie pod wpływem technologii cyfrowych. Wprowadzamy perspektywę obywatelskiej odpowiedzialności i dbania o otwartość i dobro wspólne w publicznej debacie na temat technologii.

Przykłady publikacji Centrum Cyfrowego:

A Vision for a Shared Digital Europe

Ten raport jest podsumowaniem procesu przeprowadzonego przez Kennisland, Centrum Cyfrowe i Commons Network, mającego na celu opracowania nowej wizji cyfrowych polityk publicznych w Unii Europejskiej. Autorzy wizji wyszli z założenia, że rynkowe podejście do polityk cyfrowych (jakiego byliśmy świadkami przy okazji Digital Single Market) nie jest jedynym i potrzeba perspektywy stawiającej w centrum potrzeby społeczeństwa.

Trend analysis, future applications and economics of TDM

Eksploracja tekstu i danych (TDM) ma do odegrania kluczową rolę w rozwoju Big Data. Organizacje prywatne i publiczne są coraz bardziej zainteresowane zrozumieniem, w jaki sposób podejmować decyzje w oparciu o wystarczające i rzetelne informacje, aby odnieść sukces w globalnej gospodarce opartej na wiedzy. Raport ten analizuje ekonomiczne aspekty TDM.

Twórca, buntownik, strażnik, nieświadomy użytkownik. Nauczyciele i nowoczesne praktyki edukacyjne

O tym, jak nauczyciel może używać, dzielić się lub remiksować materiały wykonane przez kogoś innego decyduje prawo autorskie. Czy ma ono wpływ na codzienną pracę nauczycieli, rozwój innowacji i kreatywność w szkole? Postanowiliśmy to sprawdzić.

Klub Jagielloński

Jesteśmy republikańskim i niepartyjnym stowarzyszeniem, które próbuje oddziaływać na politykę w duchu troski o dobro wspólne. Piszemy pogłębione artykuły o polityce, gospodarce, historii i kulturze. Formułujemy obywatelskie postulaty zmian i wysyłamy petycje do władz. Działamy w całej Polsce. Zachęcamy do odwiedzin naszego portalu publicystycznego: www.klubjagiellonski.pl.

Koncepcja raportu i koordynacja wydania były możliwe **dzięki wsparciu darczyńców Klubu Jagiellońskiego**. Jeżeli doceniasz naszą pracę, dołącz do nich! Możesz to zrobić za pomocą przelewu (PL47 1020 2892 0000 5102 0582 9454, tytułem „Darowizna na cele statutowe”), BLIKa, karty płatniczej lub przekazania 1%. Wejdź na: www.klubjagiellonski.pl/wesprzyj-nas **Dziękujemy za Twoje wsparcie!**

Think-tank Centrum Analiz Klubu Jagiellońskiego podejmuje współpracę z podmiotami zewnętrznymi w zakresie rzetelnego opracowania zagadnień wskazanych przez instytucje publiczne oraz przedsiębiorców. Podejmowane przez nas działania badawcze oraz tworzone ekspertyzy opierają się na pracy kolegialnej. Łączymy doświadczenia przedstawicieli nauki, administracji oraz biznesu.

Jednocześnie informujemy, że wszelkie działania we współpracy z podmiotami zewnętrznymi muszą być zgodne z „5 zasadami niezależności eksperckiej Centrum Analiz Klubu Jagiellońskiego”:

1. Misyjność

Klub Jagielloński, w tym Centrum Analiz Klubu Jagiellońskiego, prowadzi działalność na rzecz zmiany prawa poprzez uczestnictwo w konsultacjach publicznych, udział w posiedzeniach komisji Sejmu i Senatu, kierowanie petycji do organów władzy publicznej oraz udział w innych formach konsultowania rozwiązań prawnych organizowanych przez instytucje publiczne, polityków i organizacje obywatelskie. Działalność ta ma charakter misyjny, jest prowadzona zgodnie z zasadą prymatu dobra wspólnego i finansowana ze środków własnych stowarzyszenia. Zdecydowana większość publikowanych przez nas materiałów ma charakter misyjny. Każdorazowo, jeżeli materiał powstał we współpracy z partnerem zewnętrznym lub został sfinansowany ze środków innych niż środki własne Klubu Jagiellońskiego, jest to wyraźnie zaznaczone w jego treści.

2. Niezależność

Warunkiem podjęcia współpracy doradczej z podmiotami rynkowymi jest zagwarantowana w umowie pomiędzy stronami niezależność badawcza. Nie dopuszczamy sytuacji, w której publikujemy materiał lub opracowujemy analizę pod tezy opracowane „na zewnątrz”.

3. Przejrzystość

Warunkiem podjęcia współpracy doradczej z podmiotami rynkowymi jest zagwarantowanie w umowie pomiędzy stronami pełnej transparentności zasad współpracy. Opracowany we współpracy z zewnętrznym partnerem materiał oznaczony jest na stronie tytułowej nazwą i logo partnera. Informacja o partnerstwie zawarta jest również w ewentualnych materiałach dla mediów.

4. Samoograniczenie

Podjęcie przez stowarzyszenie lub należącą do niego spółkę współpracy z podmiotami rynkowymi w zakresie opracowania analiz, ekspertyz, raportów czy projektów aktów prawnych wyklucza dalsze zaangażowanie Klubu Jagiellońskiego w proces stanowienia prawa w obszarze bezpośrednio związanym z tematem współpracy na co najmniej 12 miesięcy od publikacji materiału powstałego w partnerstwie. Eksperti CAKJ, którzy uczestniczą w procesie stanowienia prawa, a za pośrednictwem CA lub indywidualnie wykonywali zlecenia na zamówienie podmiotów rynkowych, zobowiązani są do uprzedniego, każdorazowego i publicznego przedstawienia tej informacji (np. w stanowisku wysyłanym w ramach konsultacji publicznych lub w informacji do władz odpowiedniej komisji parlamentarnej).

5. Odpowiedzialność

W ramach współpracy z partnerami rynkowymi CAKJ nie pośredniczy w kontaktach pomiędzy ich przedstawicielami a politykami i urzędnikami. Do udziału w spotkaniach promocyjnych i dyskusyjnych wokół materiałów powstałych w partnerstwie z podmiotami rynkowymi, a zawierających rekomendacje legislacyjne, CAKJ nie zaprasza imiennie zawodowych polityków ani urzędników. Klub Jagielloński, w ramach działalności statutowej, organizuje różnego rodzaju debaty, spotkania i seminaria z udziałem przedstawicieli różnych obszarów życia publicznego, w tym osób ze świata biznesu i polityki. Spotkania tego rodzaju, gdy dotyczą zagadnień regulacyjnych i pokrewnych, zawsze mają charakter misyjny i finansowane są wyłącznie ze środków własnych stowarzyszenia.

Kontakt w sprawie współpracy: Bartosz Paszcza, członek zarządu Klubu Jagiellońskiego, bartosz.paszcza@klubjagiellonski.pl, +48698477451

