

KUBA PIWOWAR

# CO WIDZĄ ALGORYTMY?

## KONSEKWENCJE ALGORYTMICZNEJ (NIE)WIDOCZNOŚCI I (NIE)WIDZIALNOŚCI DANYCH

### KUBA PIWOWAR

Socjolog, analityk, doktorant na Uniwersytecie SWPS w Warszawie, gdzie w Katedrze Kulturoznawstwa pracuje nad rozprawą doktorską dotyczącą wykluczenia algorytmicznego. Członek Polskiego Towarzystwa Badaczy Rynku i Opinii, współorganizator kilku konferencji na styku biznesu i technologii. Pracuje w Google.

Coraz większe zrozumienie dla roli, jaką algorytmy odgrywają w społeczeństwie, sprawia, że z uwagą się im przyglądamy, poddajemy je analizie i sprawdzamy ich działanie. Zainteresowanie budzą jednak nie tylko same algorytmy, ale także dane, które są podstawą ich funkcjonowania. Szczególnie ciekawe są zagadnienia dotyczące dostępności danych wykorzystywanych przez algorytmy, ale również politycznych, społecznych, kulturowych i biznesowych sił, które nadzorują ich zbieranie i przepływ. Z tej perspektywy w niniejszym tekście przyglądam się, w jaki sposób rozmaite typy danych są wykorzystywane lub pomijane w procesie ich algorytmicznego przetwarzania. Ciekawe jest bowiem zarówno to, co dla algorytmów jest widoczne, jak i to, co umyka ich uwadze.

Algorytm jest instrukcją postępowania, która potrzebuje danych wejściowych (*input*), aby je przetworzyć i wyprodukować dane wyjściowe (*output*). W tej szerokiej definicji mieści się wiele sposobów działania: od pieczenia szarlotki<sup>1</sup>, po niektóre

---

<sup>1</sup> Z uznaniem przepisu kulinarnego za algorytm nie zgadza się na przykład Pedro Domingos. Tłumaczy to tym, że algorytm potrzebuje bardzo jasnych instrukcji, które wraz z konkretną operacjonalizacją pojęć warunkują jego prawidłowe działanie. Tym samym pieczenie ciasta nie jest algorytmem, ponieważ komputer nie poradziłby sobie z określeniami miar wielkości, takimi jak „szczypta” lub „nieco”, abstrakcyjne jest dla niego także pojęcie cukru lub jabłka. Domingos ogranicza jednak pojęcie algorytmu do nauk komputerowych (w których podnoszone przez niego zastrzeżenie o konieczności bycia precyzyjnym i niedwuznacznym jest jak najbardziej zasadne), poza którymi

aspekty zarządzania miastem. Oczywiście mniej lub bardziej proste czynności, które wpisują się w tak zarysowaną definicję algorytmu, podatne są na błędy czy niedopatrzenia lub zależą od gustu. Można upiec dobrą szarlotkę (*output*) nawet wtedy, gdy kucharz przesadzi z ilością jabłek (*input*); szarlotka wtedy jest dobra, kiedy jej konsument tak uzna, nawet jeśli jakiegoś składnika użyto w nadmiarze. Miasto będzie działać, pomimo opóźnień lub niejasnych instrukcji (algorytmów) wysyłanych między departamentami. Złożone struktury społeczne wytwarzają bufony i hamulce, za pomocą których niwelowane są błędy i niedopatrzenia. Zostały zaprojektowane na wypadek zdarzenia nieprzewidywalnego albo kolektywnie akceptowalnego „machnięcia ręką”, gdy coś dzieje się niezgodnie z planem.

Jeśli jednak algorytmy zaczniemy rozpatrywać z perspektywy technologii, wtedy będą one „zakodowanymi procedurami, które przekształcają dane wejściowe w pożądany rezultat, w oparciu o konkretne obliczenia”<sup>2</sup>. Doprecyzowanie dotyczy dwóch obszarów. Po pierwsze, w zdaniu tym stwierdzono konieczność istnienia „pożądanego rezultatu”, czyli czegoś, co można uznać za „sukces”<sup>3</sup>. Chociaż algorytm może nie zadziałać lub wynikiem jego działania może być błąd czy zapętlenie, to tylko wtedy, kiedy uznamy, że wykonał swoje zadanie zgodnie z naszymi oczekiwaniami, możemy stwierdzić, że uzyskaliśmy zamierzony rezultat. Po drugie, definicja ta wprowadza kontekst obliczeniowy, matematyczną procedurę manipulacji na liczbach zapisanych najpewniej w formie danych, cyfrowych śladów. To ważne rozróżnienie, ponieważ określa precyzyjniej pole manewru: nie każde obliczenie musi zakończyć się sukcesem, ale każda definicja sukcesu – żeby mogła zostać przetworzona przez algorytm – musi być policzalna. Społeczeństwa jednak same mogą sprowadzać swoje instytucje do policzalnych elementów. „Obliczenia [zauważa Dominique Cardon – przyp. KP] konstruuja naszą rzeczywistość, organizują i kierują nią. [...] [Jednocześnie] obliczenia możliwe są do przeprowadzenia tylko w tych społeczeństwach, które dokonały wyboru bycia policzalnymi i mierzalnymi”<sup>4</sup>.

Z punktu widzenia niniejszego tekstu konieczne jest nałożenie na tak rozumiany algorytm płaszczyzny społeczno-kulturowej, to znaczy przyjrzenie się takiemu działaniu algorytmu, które jest ściśle powiązane ze społecznym kontekstem jego użycia: zarówno z punktu widzenia danych, których źródłem są ludzie oraz ich

---

szeroka definicja tego pojęcia pozwala na objęcie nią szeregu zjawisk: opisujemy wtedy ich charakter czy sposób działania zgodny z instrukcją. Skorzystam jednak z zawężenia zaproponowanego przez Domingosa, gdyż właśnie z tak wąskiej, technologicznej definicji pojęcia algorytmu wynika szereg komplikacji społecznych (por. P. Domingos, *The Master Algorithm. How the Quest for the Ultimate Learning Machine Will Remake Our World*, Basic Books, New York 2015).

2 T. Gillespie, *Algorithm (Digital Keywords)*, Culture Digitally, 25 czerwca 2014, <http://culturedigitally.org/2014/06/algorithm-draft-digitalkeyword/> (8 lutego 2019).

3 Takie rozumienie algorytmu, w którym istnieje u d a n y wynik jego działania, ma swoje uzasadnienie nie tylko w kontekście technologicznym. Algorytmy z reguły służą osiągnięciu jakiegoś celu, przy czym w informatyce jest on – na poziomie pojedynczej operacji – nienegocjowalny. Jeśli algorytm nie zadziała, trzeba zmienić dane, sam algorytm lub definicję sukcesu.

4 D. Cardon, *Deconstructing the algorithm. Four types of digital information calculations*, [w:] *Algorithmic Cultures. Essays on Meaning, Performance, and New Technologies*, pod red. R. Seyferta, J. Roberge’a, Routledge, London–New York 2016, s. 108.

otoczenie, jak też praktyk regulacyjnych, które są przyczyną, kwintesencją albo wręcz następstwem użycia algorytmów. Pisząc zatem o algorytmach, mam na myśli takie struktury i zbiory elementów, które łączą aspekt społeczny z technicznym (*sociotechnical assemblage*)<sup>5</sup>. Choć sam algorytm jest abstrakcyjną formułą matematyczną, która potrzebuje danych do działania, to nie spełni swojej funkcji, jeśli nie zostanie zaimplementowany do konkretnego rozwiązania technologicznego – a to z kolei musi zostać użyte, aby wykonać zadanie, do którego zostało skonfigurowane<sup>6</sup>. W uproszczeniu układ ten można przedstawić w następujący sposób: dane → algorytm → implementacja do technologii → użycie<sup>7</sup>.

W zbiorach danych odbijają się współczesne społeczeństwa, z ich nierównościami, stereotypami i uprzedzeniami. Algorytmy, które te zbiory poddają obróbkę przez sortowanie, klasyfikowanie i hierarchizowanie<sup>8</sup>, są zatem elementem struktury, w której podejmowane decyzje mają znaczenie społeczne, kulturowe, prawne i etyczne. Algorytmy wykorzystywane są do określania społecznej wartości obywateli na podstawie ich zachowań<sup>9</sup>, zarządzają naszym sposobem konsumowania kultury i obcowania z nią<sup>10</sup>. Akty prawne wymieniane między instytucjami występują w ilościach niemożliwych do przetworzenia przez żadną kancelarię – stosuje się zatem techniki uczenia maszynowego, które dokonują wstępnej selekcji dostępnego materiału<sup>11</sup>. Z perspektywy etycznej algorytmy pośredniczą w podejmowaniu decyzji albo podejmują ją autonomicznie w sprawach związanych z ludzkim zdrowiem<sup>12</sup>, wolnością<sup>13</sup>, a nawet życiem<sup>14</sup>. Rozwój algorytmów sztucznej inteligencji oraz jej szerokie zastosowanie przez sektory komercyjne, a także rządy oraz instytucje publiczne, każą zwrócić uwagę na szczególnego typu wyzwania związane z deregulacją lub szybkim rozwojem wielu obszarów życia, który spotyka się z pogłębioną refleksją etyczną<sup>15</sup>.

W tak zarysowanym kontekście chcę postawić pytanie o to, jakie dane zasila ją algorytmy, skąd one pochodzą i jakie skutki ów fakt ze sobą niesie. W tekście celowo pomijam dyskusję o rodzajach algorytmów wykorzystywanych przy podejmowaniu decyzji. Niezależnie od tego, czy posłużono się prostym algorytmem

5 T. Gillespie, *The relevance of algorithms*, [w:] *Media Technologies: Essays on Communication, Materiality, and Society*, pod red. T. Gillespiego, P. Boczkowskiego, K. Foot, MIT Press, Cambridge 2012, s. 167–194.

6 B.D. Mittelstadt, P. Allo, M. Taddeo, S. Wahter, L. Floridi, *The ethics of algorithms. Mapping the debate*, „Big Data and Society”, 1 grudnia 2016, <http://journals.sagepub.com/doi/abs/10.1177/2053951716679679> (8 lutego 2019).

7 Pomijam sytuację, w której użycie technologii w kontekście społecznym może być aktem generującym nowe dane. W tym sensie zaprezentowany model powinien mieć raczej formę koła z rozgałęzieniami.

8 Por. T. Striphas, *Algorithmic culture*, „European Journal of Cultural Studies” 4–5(18)/2015, s. 395–412.

9 R. Botsman, *Big data meets Big Brother as China moves to rate its citizens*, „Wired”, 21 października 2017, <https://www.wired.co.uk/article/chinese-government-social-credit-score-privacy-invasion> (8 lutego 2019).

10 T. Striphas, *Algorithmic culture*, dz. cyt., s. 395–412.

11 J. Kaplan, *Artificial Intelligence. What Everyone Needs to Know*, Oxford University Press, New York 2016; P. Zgrzebnicki, *Selected ethical issues in artificial intelligence, autonomous system development and large data set processing*, „Studia Humana” 63(3)/2017, s. 24–33.

12 V. Eubanks, *Automating Inequality*, St. Martin's Press, New York 2018.

13 C. O'Neil, *Weapons of Math Destruction*, Penguin Books, London 2016.

14 A. Greenfield, *Radical Technologies. The Design of Everyday Life*, Verso, London–New York 2017, s. 222–226.

15 Por. P. Zgrzebnicki, *Selected ethical issue...*, dz. cyt., s. 24–33.

sortującym, czy też wyrafinowanym algorytmem sztucznej inteligencji opartym na sieciach neuronowych, pozytywne i negatywne skutki działania są społeczne i z takiej perspektywy należy je rozpatrywać.

### PAMIĘĆ KULTUROWA A WIDOCZNOŚĆ DANYCH

Opisując dynamikę pamięci kulturowej, Aleida Assmann zaproponowała jej rozbicie na formę aktywną oraz bierną – dotyczyło to zarówno samej pamięci, jak też zapominania<sup>16</sup>. Aktywne zapominanie oznacza intencjonalne niszczenie pamięci o czymś, z kolei bierne – czynności niezamierzone lub przypadkowe, które wymazują pamięć o czymś. Pamięć z kolei w swojej aktywnej formie „zachowuje przeszłość jako teraźniejszość”<sup>17</sup> przez publiczne udostępnianie zbioru (kanonu). Pamiętanie bierne polega na archiwizowaniu, „zachowuje przeszłość jako przeszłość”<sup>18</sup>. „Do aktywnej pamięci należą między innymi dzieła sztuki, które jako takie przeznaczone są do ciągłego odczytywania, podziwiania, wystawiania na scenie, odtwarzania i komentowania”<sup>19</sup>. Z perspektywy człowieka istnienie kanonu i archiwum jest konsekwencją kuratorskich zabiegów na artefaktach kulturowych – koniecznym, gdyż nie ma takiej społeczności, która miałaby dość zasobów poznawczych do ciągłego przetwarzania całej swojej spuścizny.

Dzisiaj zbierane o nas dane są niemożliwe do ręcznego przetworzenia, zatem wykorzystujemy w tym celu maszyny. Zarządzają one zbiorami danych i na podstawie instrukcji udzielają nam dostępu do części z nich, ich zdaniem najbardziej odpowiednich. Algorytmy działają na zasadzie sortowania danych<sup>20</sup> według wartości<sup>21</sup>, z jednej strony je porządkując, a z drugiej dając asumpt do decydowania o tym, co w wyniku tej operacji jest bardziej „wartościowe” (w sensie deskrypcyjnym, a nie normatywnym), a co mniej. Oczywiście pojęcie wartości podlega negocjacjom, a ustalane jest na poziomie celu, w jakim algorytm został użyty. I tak algorytmy działające w tle serwisów społecznościowych na podstawie szeregu sygnałów decydują o tym, jakie informacje wyświetlić użytkownikom, a jakie pominąć. Może to doprowadzić do sytuacji, w której na skutek nieprzejrzystego filtrowania treści, mającego na celu utrzymanie uwagi, powstaną tak zwane bańki filtracyjne<sup>22</sup>, a więc widoczne będą tylko informacje bliskie profilowi zainteresowań, określone wcześniej na podstawie szeregu sygnałów zebranych o konkretnej osobie. Tak działają: algorytm Facebooka, systemy rekomendacyjne Netflix i Amazona, a nawet przyjacielskie polecenie. W obliczu ogromu danych

16 A. Assmann, *Między historią a pamięcią*. Antologia, pod red. M. Saryusz-Wolskiej, tłum. Z. Dziewanowska-Stefańczyk i in., Wydawnictwa UW, Warszawa 2013.

17 Tamże, s. 76.

18 Tamże.

19 Tamże, s. 77.

20 Nie wszystkie algorytmy muszą być algorytmami sortującymi, niektóre mogą być zwykłą instrukcją postępowania, na przykład gdy na skutek pojawienia się nowej porcji danych przekładają je z jednego miejsca w inne.

21 C. Sandvig, *Seeing the Sort. The Aesthetic and Industrial Defense of 'The Algorithm'*, „Media-N”, styczeń 2014, <http://median.newmediacaucus.org/art-infrastructures-information/seeing-the-sort-the-aesthetic-and-industrial-defense-of-the-algorithm/> (8 lutego 2019).

22 E. Pariser, *The Filter Bubble. What the Internet is Hiding from You*, Penguin Books, London 2011.

chętnie sięgamy po pomoc tych, którzy je już przetworzyli, żebyśmy sami nie musieli tego robić. W tym sensie możemy mówić o algorytmicznej widzialności (*algorithmic visibility*)<sup>23</sup>, w rozumieniu udostępniania, uwidaczniania konkretnych treści użytkownikom, aby zrealizować założenia konkretnego reżimu widoczności (*visibility regime*)<sup>24</sup>.

Sposób zbierania, przechowywania i przetwarzania danych przypomina strukturę pamięci kulturowej: inne instytucje oraz siły regulują *Big Data* i pamięć kulturową, inne są też nośniki tej pamięci oraz ich trwałość, ale funkcja przechowywania w celu późniejszego wydobycia i przetworzenia pozostaje ta sama. Przywołuję w tym miejscu model Assmann ze względu na jego poręczność przy określaniu tego, kogo i co widzą algorytmy. Jeśli ich praca – polegająca z jednej strony na próbie odwzorowania rzeczywistości za pomocą modeli matematycznych<sup>25</sup>, a z drugiej na poszukiwaniu wzorców także w miejscach nieintuicyjnych dla człowieka obserwatora<sup>26</sup> – opiera się na ciągłym skupianiu i rozpraszaniu uwagi, wydobywaniu na powierzchnię i spychaniu poza obszar widzialności, to dostajemy narzędzie analityczne, które wskazać może nie tylko obszary potencjalnych wykluczeń, ale także miejsca poszukiwania środków zaradczych.

## WIDOCZNOŚĆ/WIDZIALNOŚĆ

Algorytm nie działa bez danych, zatem są one warunkiem *sine qua non* jego funkcjonowania. Wiele z nich jest jednak obecnych, ale niekoniecznie dostępnych. Są dane, które zostały utracone bezpowrotnie, ale też takie, które odzyskano, choć tylko częściowo. Dane się zbiera, zarządza nimi, przygotowuje, analizuje, wizualizuje – działania te wykonywane są zawsze w określonym celu, zaspokajającym czyjś interes. W tym sensie, jak zauważa Lisa Gitelman, nigdy nie są surowe, zawsze w jakiś sposób są one poddane obróbce<sup>27</sup>. Ale są także dane, których nie udało się zebrać, lub też takie, które – choć zebrane – nie są wykorzystywane. Określam ów obszar (nie)wykorzystania danych jako widoczność oraz widzialność.

Choć pojęcie widoczności algorytmicznej nie jest nowe, używane było do tej pory w opisanym wcześniej kontekście bycia widzialnym dla użytkowników mediów<sup>28</sup>. Algorytmy aktywnie sterują tym, co jest widzialne, opierając się na reżimie widoczności, w którym architektura samych mediów stanowi o dostępie do treści, ale też o polityce nadzoru nad nią<sup>29</sup>. W niniejszym tekście używam słowa „widoczność” w rozumieniu dostrzegalności – w tym przypadku danych – jako koniecznego substratu w pracy algorytmu. Choć pojęcia widoczności oraz

23 Por. J.C. Magalhães, *Algorithmic visibility. Elements of new regime of visibility*, materiały konferencyjne, kwiecień 2017, [https://www.researchgate.net/publication/321245421\\_Algorithmic\\_visibility\\_-\\_Elements\\_of\\_new\\_regime\\_of\\_visibility](https://www.researchgate.net/publication/321245421_Algorithmic_visibility_-_Elements_of_new_regime_of_visibility) (8 lutego 2019).

24 Tamże.

25 Por. C. O’Neil, *Weapons...*, dz. cyt.

26 Por. A. Greenfield, *Radical Technologies...*, dz. cyt.

27 L. Gitelman, V. Jackson, *Introduction*, [w:] „Raw Data” Is an Oxymoron, pod red. L. Gitelman, MIT Press, Cambridge-London 2013, s. 1–14.

28 J.C. Magalhães, *Algorithmic visibility...*, dz. cyt.

29 Por. tamże.



widzialności są synonimiczne, proponuję wprowadzić rozróżnienie, które dobrze oddaje strukturę napięć, jakie występują na granicy między procesami zbierania danych oraz sposobów ich wykorzystywania. Widoczność jest atrybutem jakiejś rzeczy lub osoby: widoczne jest coś, co albo zostało oznaczone w celu bycia zauważonym, albo wydobyte na powierzchnię niezależnie od bycia widocznym. Jest to atrybut bierny, to znaczy niezależny od obiektu, którego jest cechą. Na przykład osoba może być widoczna na drodze, jeśli jest w odpowiedni sposób oznakowana. Nie oznacza to jednak, że zostanie zauważona. W tym sensie widzialność jest funkcją bycia zauważonym, zakłada istnienie podmiotu obserwującego, ale też obserwowanego zjawiska, jest zatem atrybutem aktywnym. Dodatkowo widzialne może być wszystko, niezależnie od tego, czy zostało oznaczone jako widoczne, istotą widzialności jest bowiem chęć, intencja dostrzeżenia czegoś, aktywny proces poszukiwania i zwracania uwagi<sup>30</sup>. Widoczność jest zatem oznaką wyrażności; ma to spore znaczenie w kontekście algorytmów, gdyż w dużym zbiorze oznaczonych danych nie wszystkie dane muszą być widoczne, ponieważ pojedyncze elementy analizowanego zbioru mogą stracić swoją wyrazistość.

Choć z punktu widzenia człowieka pojęcia widoczności oraz widzialności są uznawane za funkcję wzroku, w przypadku algorytmów będę miał na myśli faktyczną obecność danych dowolnego rodzaju – nie tylko obrazów, ale także danych tekstowych, głosowych, metadanych<sup>31</sup> i innych. Istotny jest również fakt przekształcenia tych danych, to znaczy użycia ich przez algorytm. Dane są wtedy widoczne, kiedy zostały użyte<sup>32</sup>. Analiza (przetworzenie) danych to proces skupienia uwagi algorytmu<sup>33</sup>. Fakt ten dodatkowo powoduje, że dane – aby były widoczne – muszą zostać wywołane. Nie może być widoczne coś, co jest poza zasięgiem skupienia uwagi lub skryptu, który nastawiony jest na wydobycie jakiejś informacji. Widoczność zatem w przypadku algorytmów jest atrybutem danych, które będą przetworzone. Podmiotem działającym jest w tej sytuacji algorytm – reguluje przepływ danych, które są dla niego widzialne, choć od ich jakości oraz (nie)obecności uzależniona będzie zarówno definicja sukcesu algorytmu, jak i społecznych oraz kulturowych, zamierzonych i niezamierzonych konsekwencji jego działania<sup>34</sup>.

30 W tym samym sensie jakaś rzecz, cecha, postać – widoczna lub nie – może zostać celowo pominięta.

31 Na przykład w przypadku zdjęcia metadaną może być tag określający położenie geograficzne fotografa w momencie jego robienia, parametry techniczne lub indukowane algorytmiczne określenie tego, co lub kto się na nim znajduje.

32 Wracając do przykładu z widocznością osoby na drodze, to widoczność jest istotna, kiedy zachodzi potrzeba zwrócenia uwagi kierowców na jej obecność.

33 D. Danks, A.J. London, *Algorithmic bias in autonomous systems*, „Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI 2017)” 2017, s. 4692–4697.

34 W podobnym sensie, kiedy mówimy o stronniczości albo uprzedzeniu algorytmów (*algorithmic bias*), mamy na myśli nie tyle stronniczość samego algorytmu, ile danych wejściowych i wyjściowych, które w procesie algorytmizacji są używane (por. B.D. Mittelstadt, P. Allo, M. Taddeo, S. Wahter, L. Floridi, *The ethics...*, dz. cyt.; D. Danks, A.J. London, *Algorithmic bias...*, dz. cyt.). Sam algorytm oczywiście również może być stronniczy (*biased*), a przyczyną tego stanu rzeczy jest intencjonalne skrzywienie algorytmu w celu osiągnięcia konkretnego efektu. Tym efektem może być na przykład zniwelowanie krzywdzącej niereprezentatywności danych w taki sposób, aby zapobiec ewentualnym szkodom wymierzonym w grupy nieuprzywilejowane, na przykład kobiety, mniejszości seksualne, inne rasy (por. tamże).

Rozróżnienie na dane, które zostały zebrane, oraz takie, które zostały przetworzone przez algorytm, oznacza jednocześnie istnienie takich danych, które nie zostały zebrane, ale także takich, które – choć zebrane – nie zostały użyte. Ten drugi typ danych można uznać za niewidoczne – chwilowo lub systemowo. Opisane typy danych przedstawiam w tabeli.

Tabela. Widoczność i widzialność danych

6 Dane niezbrane			
5 Dane zebrane, ale nieużywane lub niedostępne			
<b>Dane zebrane</b>			
1 Widoczne		3 Niewidoczne	
2		4	
Widzialne	Niewidzialne	Widzialne	Niewidzialne
Algorytmy			

W miejscach styku między widocznością a widzialnością zachodzić mogą ciekawe procesy: (1) dane widoczne i widzialne są w zasadzie podstawą działania algorytmów, są cyfrowym kanonem stanowiącym o analizowanej materii (na przykład dane o wyszukiwaniach i ich wykorzystaniu w internetowej reklamie); (2) dane widoczne, ale niewidzialne zostały oznaczone do ewentualnego wykorzystania w pracy algorytmów, ale z różnych przyczyn nie zostały dostrzeżone i użyte (powstaje ważne pytanie, dlaczego tak się stało); (3) dane niewidoczne, ale widzialne z perspektywy algorytmu nabierają znaczenia w całej swojej masie, a nie w poszczególnych bitach informacji; (4) dane niewidoczne i niewidzialne, choć zostały zebrane i archiwizowane, są w sposób systematyczny pomijane (na przykład kiedy algorytm celowo pomija jakieś grupy w analizie).

Dodatkowo warto zwrócić uwagę, że istnieją dane zebrane i przechowywane na rozmaitych nośnikach niedostępnych algorytmom: są to przede wszystkim dane niezapisane w formie cyfrowej (5). W końcu istotne jest wspomnienie o danych, których nie zebrano (6) – ważne z punktu widzenia niniejszego tekstu jest nie tyle określenie rodzaju danych, jakich nie zebrano (co byłoby zadaniem dość karkołomnym), ile diagnoza przyczyn tego stanu rzeczy.

## CO WIDZĄ ALGORYTMY?

### 1. KONTEKST

W dyskusji o wielkich zbiorach danych *Big Data* podkreślanych jest kilka ich atrybutów. Pierwszy i bodaj najważniejszy, dający asumpt do nazwania samego zjawiska, mówi o wielkości danych (*volume*). Po drugie, dane zbierane są szybko, aby na ich podstawie podjąć decyzje, cechą *Big Data* jest zatem szybkość ich

zbierania (*velocity*). Po trzecie, podkreśla się ich różnorodność (*variety*), to znaczy dane zbierane są w różnych celach i z różnych źródeł. Niektórzy badacze i badaczki wskazują ponadto na kilka innych cech wielkich zbiorów danych: ich wiarygodność (*veracity*), wartość (*value*), wirtualność (*virtual*) i zmienność (*variability*)<sup>35</sup>. Pierwsze trzy „v”<sup>36</sup> przypisywano *Big Data* od początku, podczas gdy kolejne pojawiały się wraz z refleksją nad tym zjawiskiem. Kate Crawford i dannah boyd<sup>37</sup> zwracają uwagę na kilka zjawisk, które towarzyszą rozwojowi *Big Data*, podając w wątpliwość między innymi obiektywność i dokładność wielkich zbiorów danych („twierdzenie o obiektywności i dokładności są zwodnicze”), ich wartość („więcej danych nie musi oznaczać lepszych danych”) lub sam fakt wykorzystywania ich do analizy („tylko dlatego, że coś jest dostępne, nie sprawia, że wykorzystanie tego jest etyczne”). Zatem, choć rozwój zjawiska stwarza niewątpliwie szanse dla nauki lub biznesu, to jednocześnie budzi zrozumiałe obawy, także ze względu na fakt, że wykorzystywane jest ono do realizacji coraz większej liczby celów.

Dużą mocą obliczeniową oraz danymi, na których te operacje są wykonywane, posługują się zarówno sektor publiczny, jak i prywatny. Stosowane są w technologii, rolnictwie, reklamie, rozrywce, finansach, służbie zdrowia; regulują prace rządów, wspomagają między innymi handel, przemysł i górnictwo<sup>38</sup>. W ostatnich latach szczególnym zainteresowaniem cieszy się zastosowanie danych w rozwoju sztucznej inteligencji. McKinsey wskazuje na pięć dziedzin tego rozwoju: technologie rozpoznawania i przetwarzania obrazów, technologie przetwarzania języka, wirtualni asystenci, autonomiczne roboty i pojazdy oraz uczenie maszynowe<sup>39</sup>. W dyskusji o sposobach używania sztucznej inteligencji, oprócz masowo wykorzystywanych produktów i rozwiązań, ważne jest też to, czy jest możliwe spełnienie obietnic, które składa ta technologia. W związku z pracami nad autonomicznymi pojazdami zwraca się uwagę nie tylko na prawno-regulacyjny charakter tego wyzwania, ale także na jak najbardziej aktualne problemy i pytania, które rodzi rozwój takiej technologii. Przez pryzmat problemów, o których piszą badacze i badaczki zajmujący się analizą pracy algorytmów, zwracam uwagę na fakt, że – jak wskazują Crawford i boyd<sup>40</sup> – sama obecność danych nie jest wystarczającym powodem do rozpoczęcia analizy. Konieczne jest bowiem pracowanie na wartościowych, zróżnicowanych, dobrych jakościowo zbiorach, ze szczególnym uwzględnieniem ich potencjalnych ograniczeń. Mam na myśli nadmiar lub

35 Por. P. Idzik, *Analiza Big Data. Badania niereaktywne w erze Internetu 2.0*, [w:] *Zwrot cyfrowy w humanistyce*, pod red. A. Radomskiego, R. Bomby, E-naukowiec, Lublin 2013, s. 153–168.

36 Od angielskich nazw cech.

37 d. boyd, K. Crawford, *Six provocations for Big Data*, „A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society”, wrzesień 2011, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1926431](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431) (8 lutego 2019).

38 M. Belfiore, *How 10 industries are using big data to win big*, IBM, 26 lipca 2016, <https://www.ibm.com/blogs/watson/2016/07/10-industries-using-big-data-win-big/> (8 lutego 2019).

39 *Revolucja AI: jak sztuczna inteligencja zmieni biznes w Polsce*, McKinsey & Company, 2017, <https://mckinsey.pl/publikacje/raport-rewolucja-ai-jak-sztuczna-inteligencja-zmieni-biznes-w-polsce/> (8 lutego 2019).

40 d. boyd, K. Crawford, *Six provocations...*, dz. cyt.



niedobór danych, nieoczywiste połączenia między nimi, błędy w ocenie sytuacji, ale także znaczące nieobecności. Szczególnym typem danych są tzw. *proxies*, które – z braku dobrego polskiego odpowiednika tego słowa – pozostawiam w angielskiej wersji. *Proxy* oznacza coś pośredniego, punkt zaczepienia najbliższy celowi pomiaru, ale nieoddający w pełni jego sensu. Dobrym przykładem jest miejsce urodzenia jako potencjalny wskaźnik tego, jakim językiem posługuje się dana osoba jako językiem ojczystym. *Proxies* będą się nieraz przewijać przez wskazane obszary widoczności danych.

## 2. NADMIAR I NIEDOBÓR DANYCH: DANE NIEWIDOCZNE I NIEWIDZIALNE

Joy Buolamwini i Timnit Gebru, badaczki z MIT, zwracają uwagę, że choć liczba danych użytych w aplikacjach do rozpoznawania twarzy jest ogromna, to nie są one wystarczająco dobre jakościowo. Korzystanie z nich może wykluczać niektóre grupy, ponieważ w pakietach, na których te aplikacje się uczą, nie ma na ich temat żadnych danych. Buolamwini i Gebru sprawdziły trzy komercyjnie dostępne aplikacje do rozpoznawania twarzy i udowodniły, że najlepiej radzą sobie one z twarzami białych mężczyzn, gorzej z twarzami osób o ciemnej karnacji, natomiast najgorsze wyniki uzyskują, kiedy stawia się przed nimi zadanie rozpoznania czarnoskórej kobiety<sup>41</sup>. Przyczyną takiego stanu rzeczy jest fakt, że aktualnie dostępne bazy danych budowane są na podstawie zbiorów twarzy o jasnym odcieniu. Widoczność wynikająca z nadmiaru białych twarzy jest jednocześnie przyczyną osłabienia widoczności twarzy o ciemnej skórze. Zwracano na to uwagę już w latach sześćdziesiątych i siedemdziesiątych ubiegłego wieku. Przemysł fotograficzny używał tzw. *Shirley cards* do kalibracji koloru na zdjęciach – punktem odniesienia były fotografie białych kobiet o długich włosach. Osoby o ciemnej skórze wychodziły na takich zdjęciach niewyraźnie. Uwagę na to zwrócił także przemysł meblarski: w drukowanych reklamach jego produktów odwzorowanie spektrum brązu okazało się niewystarczające<sup>42</sup>.

W przypadku projektu Buolamwini i Gebru zaproponowano wiele rozwiązań. Microsoft, którego aplikację badaczki sprawdzały, poprawił dane, na których aplikacja uczy się rozpoznawania twarzy, rozpoczął zbieranie nowych informacji z uwzględnieniem różnic płci, wieku oraz koloru skóry oraz poprawił sam algorytm<sup>43</sup>. Na problem z ograniczoną różnorodnością danych zwracają też uwagę między innymi przedsiębiorstwa w Chinach, w których rozwój prac nad sztuczną inteligencją nabiera rozpędu<sup>44</sup>. We współpracy z rządem Zimbabwe firma CloudWalk z Kantonu wprowadzi swoje rozwiązania służące rozpoznawaniu

41 J. Buolamwini, T. Gebru, *Gender shades. Intersectional accuracy disparities in commercial gender classification*, „Proceedings of Machine Learning Research” 81/2018, s. 1–15.

42 J. Lovejoy, *Fair is not the default*, Google Design, 15 lutego 2018, <https://design.google/library/fair-not-default/> (8 lutego 2019).

43 J. Roach, *Microsoft improves facial recognition technology to perform well across all skin tones, genders*, Microsoft, 26 czerwca 2018, <https://blogs.microsoft.com/ai/gender-skin-tone-facial-recognition-improvement/> (8 lutego 2019).

44 J. Roachel, C. O’Neil, *Guest post: China’s social credit system*, Mathbabe, 2 października 2018, [www.mathbabe.org/2018/10/02/guest-post-chinas-social-credit-system/](http://www.mathbabe.org/2018/10/02/guest-post-chinas-social-credit-system/) (8 lutego 2019).

twarzy, które będą wykorzystywane w rozmaitych programach publicznych związanych między innymi z bezpieczeństwem oraz przestrzeganiem prawa<sup>45</sup>.

Nadmiar danych może też przyczynić się do takiego „skrzywienia” pracy algorytmu, w którym na skutek zbyt wielkiej liczby dostępnych sygnałów pewne grupy będą poddawane większemu nadzorowi. Cathy O’Neil<sup>46</sup> zwróciła uwagę na zjawisko samospełniającego się proroctwa w sytuacji, w której algorytm sugerujący prawdopodobne miejsce popełnienia przestępstwa operuje na danych dotyczących zgłoszeń przestępstw oraz zatrzymań. To bardzo istotne, dlatego że podstawą pracy algorytmu, czyli aktywnie widocznym typem danych, są dane o zgłoszonych przestępstwach oraz zatrzymaniach, a nie wszystkich możliwych. Istnieje zatem prawdopodobieństwo, że pewne miejsca będą szczególnie widoczne w systemach policyjnych, a tym samym narażone na powtórne wskazanie przez algorytm.

### 3. NIEOCZYWISTE POŁĄCZENIA: DANE NIEWIDOCZNE, ALE WIDZIALNE

Według agencji badawczych Gartner i IDC nawet 80 procent danych to tzw. ciemne dane (*dark data*), a więc zebrane, ale nieustrukturyzowane i trudno dostępne, także takie, które nie zostały podłączone do globalnej sieci wymiany (na przykład dane analogowe lub cyfrowe, ale przechowywane na starych dyskach)<sup>47</sup>. Potencjalność ich użycia i zarchiwizowanie – ustrukturyzowane lub nie – sprawia, że charakteryzują się bierną widocznością. Dane te mogą jednak zostać użyte, mimo że nie są konkretnie wskazane do tego użycia, na przykład w sytuacji, w której algorytmy uczenia maszynowego wykorzystują zasoby danych w poszukiwaniu wzorców, także tych mniej oczywistych.

Wśród opublikowanych na przestrzeni ostatnich lat badań szczególną uwagę zwraca kilka głośnych projektów przeprowadzonych przez badaczy i badaczki w obszarze widzenia komputerowego (*computer vision*). W doświadczeniu przeprowadzonym przez Michała Kosińskiego oraz Yilun Wanga z Uniwersytetu Stanforda wykorzystano głębokie sieci neuronowe – jedną z metod sztucznej inteligencji – do określenia orientacji seksualnej na podstawie zdjęcia twarzy. Skorzystano w tym celu z ogólnodostępnej aplikacji do rozpoznawania twarzy, ale wcześniej nauczono ją rozpoznawać twarz na podstawie konkretnych tagów, czyli oznaczeń. Zastosowano więc sposób, który w uczeniu maszynowym nazywa się nadzorowanym. Tagiem była orientacja seksualna osoby na zdjęciu, a same zdjęcia pochodziły z portalu randkowego i Facebooka.

*Przy pojedynczym zdjęciu klasyfikator prawidłowo odróżnił osoby homoseksualne od heteroseksualnych w 81% przypadków u mężczyzn oraz w przypadku 71% kobiet. Ludziom*

<sup>45</sup> L. Chutel, *China is exporting facial recognition software to Africa, expanding its vast database*, Quartz, 25 maja 2018, [www.qz.com/africa/1287675/china-is-exporting-facial-recognition-to-africa-ensuring-ai-dominance-through-diversity/](http://www.qz.com/africa/1287675/china-is-exporting-facial-recognition-to-africa-ensuring-ai-dominance-through-diversity/) (8 lutego 2019).

<sup>46</sup> C. O’Neil, *Weapons...*, dz. cyt.

<sup>47</sup> M. Kanellos, *The five different types of Big Data*, „Forbes”, 11 marca 2016, <https://www.forbes.com/sites/michaelkanellos/2016/03/11/the-five-different-types-of-big-data/#1ad4a06d6f87> (8 lutego 2019).

oceniającym zdjęcia udało się osiągnąć znacznie niższy wynik: 61% dla mężczyzn oraz 54% dla kobiet. Dokładność algorytmu wzrasta, odpowiednio do 91% i 83%, jeśli [algorytm – przyp. KP] ocenia orientację seksualną osoby na podstawie pięciu zdjęć twarzy<sup>48</sup>.

Badanie to zwraca uwagę na niebezpieczeństwa, jakie niesie ze sobą łączenie różnych typów danych: wyciąganie wniosków o orientacji seksualnej z rysów twarzy budzi niebezpieczne skojarzenia z fizjonomiką, chociaż autorzy badania się od niej odzegnują, ważne są też problemy związane ze społecznym odbiorem takich działań. Dlatego Kosiński i Wang podkreślają, że istotnym elementem komunikowania wyników badań jest jednoczesna edukacja ich odbiorców, aby badanie, które wykorzystuje pracujące na prawdopodobieństwach sieci neuronowe, nie zostało użyte jako predyktor czyjejs orientacji seksualnej, czy to w warunkach domowych, czy to w ramach państwowej inwigilacji. W końcu badanie pokazuje, że algorytmy – jeśli odpowiednio je ukierunkujemy – uczynią widocznymi dane niekoniecznie będące dobrym predyktorem (*proxy*) cechy, którą mają mierzyć. W prostym eksperymencie sprawdzającym założenia poczynione przez Kosińskiego i Wanga grupa badaczy podważyła ich zasadność. Dowiedli oni, że analizowane cechy twarzy są *de facto* pochodną stylu życia i sposobów autoprezentacji i nie są wrodzone<sup>49</sup>.

#### 4. BŁĘDY W OCENIE SYTUACJI: DANE WIDOCZNE, ALE NIEWIDZIALNE

Według danych Światowej Organizacji Zdrowia każdego roku na świecie ginie w wypadkach samochodowych ponad 1,2 miliona ludzi<sup>50</sup>. To jedna z motywacji do pracy nad autonomicznymi pojazdami, której celem jest ograniczenie czynnika ludzkiego i doprowadzenie tym samym do zmniejszenia liczby wypadków. Równoległe z eksperymentami prowadzonymi przez takie firmy jak Tesla, Waymo czy BMW, odbywa się zakrojona na szeroką skalę dyskusja dotycząca etyki maszyn. Wyzwaniem, przed którym staje rynek motoryzacyjny, ale też w zasadzie cała ludzkość, jest zbudowanie pojazdu nie tylko wygodnego i szybkiego, ale także takiego, który zadba o bezpieczeństwo na drodze. W jednym z badań nad tym zagadnieniem podjęto próbę zrozumienia wyborów moralnych w sytuacji nieuniknionego wypadku. Badacze postawili pytanie, w jaki sposób powinien zachować się wówczas autonomiczny pojazd. Wyniki pokazują, jak trudne jest to zagadnienie. Osoby badane zgodziły się z utylitarnym podejściem (opartym na Benthamowskim konsekwencjalizmie<sup>51</sup>), w którym w hipotetycznym scenariuszu nieuniknionego wypadku ginie pasażer, ocalając przechodniów. Problem

48 Y. Wang, M. Kosiński, PsyArXiv, *Deep neural networks are more accurate than humans at detecting sexual orientation from facial images*, 7 września 2017, <https://psyarxiv.com/hv28a/> (8 lutego 2019).

49 B. Agüera y Arcas, A. Todorov, M. Mitchell, *Do algorithms reveal sexual orientation or just expose our stereotypes?*, Medium, 11 stycznia 2018, [www.medium.com/@blaisea/do-algorithms-reveal-sexual-orientation-or-just-expose-our-stereotypes-d998fafdf477](http://www.medium.com/@blaisea/do-algorithms-reveal-sexual-orientation-or-just-expose-our-stereotypes-d998fafdf477) (8 lutego 2019).

50 *Global Health Observatory (GHO) data: Road traffic deaths*, WHO, [https://www.who.int/gho/road\\_safety/mortality/en/](https://www.who.int/gho/road_safety/mortality/en/) (8 lutego 2019).

51 E. Awad, *Moral Machine. Perception of Moral Judgment Made by Machines*, Massachusetts Institute of Technology, czerwiec 2017, <https://dam-prod.media.mit.edu/x/2017/06/18/awad-ms-17.pdf> (8 lutego 2019).

polega na tym, że te same osoby nie chciałyby kupić samochodu, którego wybory w sytuacji opisanej w badaniu byłyby właśnie takie<sup>52</sup>. Badanie to rozwinięto w zakrojonym na szeroką skalę eksperymencie, który wykorzystał w tym celu platformę internetową Moral Machine. W rezultacie otrzymano 30 milionów odpowiedzi od ponad 3 milionów ludzi i rozbudowano nieco scenariusze wypadku o dodatkowe elementy, na przykład przez wprowadzenie jako potencjalne ofiary osób o konkretnych cechach (kobiet, ludzi starszych, osób łamiących prawo), ale też zwierząt domowych<sup>53</sup>. Choć wyniki badania nie wniosły wiele do naszego rozumienia moralności oraz tego, jak można ją zastosować w autonomicznych pojazdach, to stanowią kolejny ważny krok w poszukiwaniach takich rozwiązań oraz, w moim przekonaniu, osławiają ludzi z technologią, która może być zastosowana w przyszłości. Nie budząc strachu, powoli buduje się zaufanie.

Bywa ono jednak nadwyreżane, jak w sytuacji wypadku, który miał miejsce w samochodzie Tesli. Reklamowany jako pojazd z autopilotem, został kupiony przez jednego z entuzjastów marki. Jednak jadąc po prostej drodze, wjechał on w skręcający traktor – jasny bok pojazdu złał się z niebem w tle, przez co ani kierowca, ani kamery zamontowane w samochodzie nie spostrzegły niebezpieczeństwa. Kierowca zginął<sup>54</sup>. Adam Greenfield zauważył, że problemem jest tutaj nie tylko niedoskonałość algorytmów zarządzających pracą pojazdu: choć samochody Tesli to tak naprawdę urządzenia do zbierania danych, to jest jeszcze sporo sytuacji, w których nie umieją sobie one poradzić (to znaczy takich, w których nie zebrały danych). Wciąż jesteśmy daleko od w pełni autonomicznych pojazdów. Według Greenfielda wypadek miałby szansę się nie wydarzyć, gdyby nie nazwa produktu: słowo „autopilot” sugeruje, że kierowca może zdjąć ręce z kierownicy i pozwolić na w pełni samodzielną jazdę auta. Tymczasem nie istnieją – a przynajmniej nie zostały dopuszczone do powszechnego użytku – algorytmy, które mogłyby całkowicie zdjąć z kierowcy odpowiedzialność za prowadzenie pojazdu. Opisana sytuacja zwraca uwagę na fakt, jak bardzo potrzebne są zarówno edukacja poprzedzająca wprowadzenie konkretnych rozwiązań, jak i odpowiedzialne zarządzanie wyobraźnią opinii publicznej. „Żaden z tych [algorytmów – przyp. KP] nie jest prosty sam w sobie. [...] Natomiast rozwijanie podstawowego wyobrażenia na temat tego, co one robią, pozwala krytycznie przyjrzeć się umowie, na jaką się zgadzamy, gdy oddajemy kontrolę nad sytuacją ocenie algorytmów” – zauważa Greenfield<sup>55</sup>.

Napięcie między bezrefleksyjnym zaufaniem technologii a awersją do jej – nawet słusznych – wyborów jest ciekawym polem do badań. Zwraca uwagę na palącą potrzebę szerokiej edukacji na temat zmian technologicznych, ale także głębokiego zrozumienia przyczyn wyborów i decyzji, jakie podejmują ludzie. Z jednej strony, nie potrafimy poprawnie wskazać, w jakim celu oraz jakie dane na nasz

52 J.-F. Bonnefon, A. Shariff, I. Rahwan, *The social dilemma of autonomous vehicles*, „Science” 6293(352)/2016, s. 1573–1576.

53 E. Awad, *Moral Machine...*, dz. cyt.

54 Por. A. Greenfield, *Radical Technologies...*, dz. cyt., s. 222–226.

55 Tamże, s. 225–226.

temat są zbierane. Choć rozumiemy, że mogą być one wykorzystywane na przykład w celach marketingowych, to nie spodziewamy się, jak istotne połączenia na podstawie zebranych danych są odnajdywane<sup>56</sup>. Wnioskowanie na podstawie danych behawioralnych zbieranych w internecie wykazało wysoką skuteczność w przewidywaniu cech wrażliwych, takich jak etniczność lub poglądy polityczne<sup>57</sup>. Z drugiej strony, badacze obserwują awersję do decyzji podejmowanych przez algorytmy, w wyniku której chętniej wybieramy mniej dokładnych i częściej mylących się ludzi niż precyzyjne algorytmy. „Uczestnicy badania, którzy widzieli, w jaki sposób algorytm podejmuje decyzje, byli mniej pewni jego wyborów oraz wskazywali go mniej chętnie w porównaniu z gorzej radzącym sobie z zadaniem człowiekiem”<sup>58</sup>. Przyczyną tego stanu rzeczy były błędy popełniane przez algorytm. Jeśli były ciągle te same, uczestnicy badania tracili zaufanie do skuteczności algorytmu, co z kolei nie miało miejsca, gdy w identyczny sposób mylił się człowiek.

Ten ciekawy przykład unaocznia, w jak różny sposób dane gromadzone przez ludzi oraz algorytmy mogą być niewidoczne. Nawet odpowiednio otagowane i poszukiwane, nie są widzialne, gdy zlewają się z tłem. Możemy wręcz nie wiedzieć o ich obecności oraz wykorzystaniu – czasem bowiem zebrane o nas dane są widzialne przez algorytmy, ale niewidoczne dla nas, podobnie jak połączenia, które algorytmy między nimi tworzą, czy wnioski, które wyciągają. W końcu dane mogą zostać zignorowane – wniosek niebanalny, jeśli weźmiemy pod uwagę fakt, że nasza preferencja do wybierania ludzi, a nie maszyn, jest także punktem wyjścia w dyskusji na temat zaufania do technologii.

## 5. NIEOBECNI, WYKLUCZENI

Liczba dostępnych danych jest trudna (jeśli nie niemożliwa) do wyobrażenia. Codziennie ludzkość wytwarza tyle danych, że potrzeba by było 250 tysięcy Bibliotek Kongresu, żeby je pomieścić<sup>59</sup>. Tak zarysowany punkt odniesienia nie zmienia jednak faktu, że liczbę tę trudno skonceptualizować. Danych jest tak dużo, że Chris Anderson, onegdaj redaktor naczelny magazynu „Wired”, ogłosił „koniec teorii”, uznając, że nie musimy już budować modeli predykcyjnych – ogrom danych zebranych o ludziach mówi sam za siebie<sup>60</sup>. Czy na pewno? Crawford i Boyd zwracają uwagę, że narzędzia prowadzące analizy i badania automatycznie (czyli same z siebie) nie muszą być doskonałe, jednocześnie podkreślają, że duże zbiory

56 J. Warshaw, N. Taft, A. Woodruff, *Intuitions, analytics, and killing ants. Inference literacy of high school-educated adults in the US*, „Proceedings of the Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)”, 22–24 czerwca 2016, <https://www.usenix.org/system/files/conference/soups2016/soups2016-paper-warshaw.pdf> (8 lutego 2019).

57 M. Kosiński, D. Stillwell, T. Graepel, *Private traits and attributes are predictable from digital records of human behavior*, PNAS, 9 kwietnia 2013, <https://www.pnas.org/content/110/15/5802> (8 lutego 2019).

58 B.J. Dietvorst, J.P. Simmons, C. Massey, *Algorithm aversion. People erroneously avoid algorithms after seeing them err*, „Journal of Experimental Psychology: General” 1(144)/2015, s. 114–126, <https://psycnet.apa.org/doiLanding?doi=10.1037%2F1037-0000033> (8 lutego 2019).

59 M. Khoso, *How much data is produced every day?*, Northeastern University, 13 maja 2016, <http://www.northeastern.edu/levelblog/2016/05/13/how-much-data-produced-every-day/> (8 lutego 2019).

60 Ch. Anderson, *The end of theory: the data deluge makes the scientific method obsolete*, „Wired”, 23 czerwca 2008, <https://www.wired.com/2008/06/pb-theory/> (8 lutego 2019).



danych nie muszą oznaczać, że ich jakość jest wystarczająco wysoka, aby użyć ich w wartościowych procesach produkowania wiedzy<sup>61</sup>. Nie bez znaczenia pozostaje fakt, że nie wszystko zostało objęte pomiarem. Jak wskazuje World Economic Forum, połowa ludzkości nie ma dostępu do internetu<sup>62</sup>. Oznacza to, że ponad 3,5 miliarda ludzi nie pozostawia cyfrowych śladów, które mogłyby zostać przetworzone przez algorytmy. Są to ludzie zamieszkujący Afrykę, kraje arabskie oraz rejon Azji i Pacyfiku<sup>63</sup>, a zatem w dużej mierze kraje rozwijające się. „Okolo 75% niepodłączonej do internetu populacji skupiona jest w 20 krajach, zamieszkuje obszary wiejskie, ma niskie dochody, jest starsza, niepiśmienna i są to przede wszystkim kobiety”<sup>64</sup>.

Źródłem danych mogą być osoby nie tylko podłączone do internetu przez komputery lub urządzenia mobilne, ale także wszelkiego rodzaju urządzenia, których działanie nabiera sensu w momencie podłączenia ich do sieci: mam tu na myśli zjawisko nazwane „internetem rzeczy” (*Internet of Things*). Agencja badawcza Gartner szacuje, że w 2017 roku na świecie było około 8,4 miliarda urządzeń podłączonych do internetu, a liczba ta ma wzrosnąć do 20,4 miliarda w roku 2020. Nie dziwi fakt, że dwie trzecie tych urządzeń trafiło w ręce konsumentów z Chin, Ameryki Północnej i zachodniej Europy<sup>65</sup>. Olbrzymie zbiorowości nie mają dostępu do nowych technologii lub dysponują nimi w bardzo ograniczonym zakresie, co wyklucza je z aktywnego uczestnictwa w tak zwanej czwartej rewolucji przemysłowej, za jaką uważa się wplecenie zmian technologicznych (szybkich i opartych na danych) w tkanę społeczną, gospodarczą, kulturową i naukową<sup>66</sup>.

Obszary niewidoczności, nieoznakowania, nieopisania olbrzymich połąci ludzkiej działalności odzwierciedlają w mniej lub bardziej oczywisty sposób linie demarkacyjne między społeczeństwami biednymi a bogatymi. Wykluczenie z pola opisu z powodu braku danych cyfrowych pochodzących z wszelkiego rodzaju urządzeń elektronicznych jest siłą rzeczy pochodną sił politycznych i gospodarczych<sup>67</sup>. Choć o biedniejszej połowie ludzkości nie wiemy tyle, co o bogatszej, zbieramy na jej temat dane i próbujemy je za pomocą rozmaitych narzędzi mierzyć<sup>68</sup>.

61 d. boyd, K. Crawford, *Six provocations...*, dz. cyt.

62 E. White, O. Pinsky, *Half the world's population is still offline. Here's why that matters*, World Economic Forum, 14 maja 2018, <https://www.weforum.org/agenda/2018/05/half-the-world-s-population-is-still-offline-heres-why-that-matters/> (8 lutego 2019).

63 Tamże.

64 *Offline and falling behind. Barriers to Internet adoption*, McKinsey & Company, wrzesień 2014, <https://www.mckinsey.com/industries/high-tech/our-insights/offline-and-falling-behind-barriers-to-internet-adoption> (8 lutego 2019).

65 R. van der Meulen, *Gartner says 8.4 billion connected „things” will be in use in 2017, up 31 percent from 2016*, Gartner, 7 lutego 2017, <https://www.gartner.com/en/newsroom/press-releases/2017-02-07-gartner-says-8-billion-connected-things-will-be-in-use-in-2017-up-31-percent-from-2016> (8 lutego 2019).

66 Tamże; K. Schwab, *The Fourth Industrial Revolution: what it means, how to respond*, World Economic Forum, 14 stycznia 2016, <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/> (8 lutego 2019).

67 Por. K. Crawford, V. Joler, *Anatomy of an AI System*, 2018, [www.anatomyof.ai](http://www.anatomyof.ai) (8 lutego 2019).

68 Wątek ten wymaga szerokiego omówienia, na które w niniejszym tekście nie ma miejsca. Chodzi o niepodłączone do internetu, biedniejsze rejony świata, które poddawane są różnym formom nadzoru ze strony państw, występujące też jako mniej interesujące grupy docelowe, będące poza działaniami reklamodawców i z tej



Nie jest to miejsce na stawianie pytań o granice pomiaru i wynikającego z niego nadzoru: czy lepiej być wykluczonym z obiegu cyfrowych śladów, ale być wolnym od inwigilacji i nadzoru, czy też zrezygnować z części wolności, ale uzyskać prawo do szczegółowego (czasami nawet za bardzo) opisu? Jednak konsekwencje braku prowadzenia statystyk i pomiarów mogą być różne i także sprzyjać wykluczeniu. Ciekawą ilustrację negatywnych skutków braku pomiaru przytacza Reni Eddo-Lodge:

*Jeden z paragrafów [ustawy o włóczęgostwie – przyp. KP] z 1824 roku dawał policji prawo zatrzymania, przeszukania i aresztowania każdego, kogo podejrzewała o możliwość popełnienia przestępstwa. Ustawa zasłynęła jako sus laws (od użytego w niej sformułowania suspected person, oznaczającego podejrzanego). Ponieważ policja nie prowadziła statystyk dotyczących zatrzymań zgodnych z ustawą, trudno określić, ile osób było zaczepianych przez policję tylko dlatego, że nie wyglądały na szacownych obywateli. Na podstawie zbieranych przez siebie dowodów działacze antyrasistowscy twierdzili, że ofiarą sus laws często padają czarni. Rasizm niewątpliwie w wielu przypadkach przesądzał o tym, kto wygląda podejrzenie, a kto nie – szczególnie w klimacie politycznym Wielkiej Brytanii, gdzie zaledwie dziesięć lat wcześniej odmawiano czarnym zatrudnienia i kwaterunku<sup>69</sup>.*

#### PODSUMOWANIE: PROXIES

Ze względu na niski koszt przechowywania *Big Data* oraz mocy obliczeniowej zbieranie danych dla samego zbierania ziszcza plan, w którym siecią cyfrowego opisu objęte zostaje w miarę możliwości jak najwięcej aspektów ludzkiej aktywności. Koncepcję kompletnego opisanie rzeczywistości w postaci cyfrowych śladów zawarł w jednym ze swoich opowiadań Jorge Luis Borges: w przedstawionym przez niego świecie myśl kartograficzna była tak rozwinięta, że zdołano odwzorować imperium przez pełne pokrycie go mapą, w skali 1:1<sup>70</sup>. Ta utopijna i w zasadzie niepraktyczna idea stała się tak naprawdę celem przyświecającym pracom nad uczeniem maszynowym: „pełną kwantyfikacją wizualnych, słuchowych i poznawczych reżimów rzeczywistości. Od kosmologicznego modelu wszechświata do świata ludzkich emocji interpretowanych przez najdrobniejsze ruchy mięśni twarzy – wszystko staje się przedmiotem kwantyfikacji<sup>71</sup>”. Jest to jednak aspiracja daleka od pełnej realizacji. Dzisiaj w zbiorach danych są luki, mniej lub bardziej znaczące ślepe punkty, które algorytmy albo pomijają, albo ich nie dostrzegają, albo też próbują je wypełnić, pokrywając je siatką prawdopodobieństw. W sytuacji braku danych i potrzeby określenia zawartości tej nieskwantyfikowanej przestrzeni budowane są modele, które z pewną dozą prawdopodobieństwa mogą temu zaradzić.

---

perspektywy badane. Mam na myśli także rozmaite obszary wykluczeń wewnątrz społeczeństw rozwiniętych, takie jak obszary bezdomności.

<sup>69</sup> R. Eddo-Lodge, *Dlaczego nie rozmawiam już z białymi o kolorze skóry*, tłum. A. Sak, Wydawnictwo Karakter, Kraków 2018, s. 54–55.

<sup>70</sup> Por. K. Crawford i V. Joler, *Anatomy of an AI System*, dz. cyt.

<sup>71</sup> Tamże.

W Stanach Zjednoczonych silnym predyktorem rasy jest kod pocztowy miejsca zamieszkania. Choć zatem w algorytmicznych systemach wykorzystywanych do selekcji miejsc, w które wysyła się patrole policyjne, nie musi być informacji o rasie, to za pomocą odpowiednio skalibrowanych modeli systemy te dochodzą do takich wniosków, jakby te dane posiadały, tym samym wzmacniając uprzedzenia, a zarazem wiarę w neutralność technologii. Jeśli algorytm nie miał danych dotyczących rasy, ale sam doszedł do wniosku, że niektóre regiony są bardziej niebezpieczne niż inne, to wzmocnił wrażenie, że widoczne dane są gwarantem obiektywności<sup>72</sup>.

Kate Crawford zwraca uwagę, że „przy okazji każdego dużego zbioru danych musimy zadać sobie pytanie, jakich ludzi wykluczaliśmy. Które obszary są mniej widoczne? Co się dzieje, jeśli żyjesz w cieniu tych zbiorów [pozostając niezauważonym – przyp. KP]?”<sup>73</sup>. Oplecenie wszystkich aspektów ludzkiej działalności danymi, sprowadzenie do policzalnych wartości takich zagadnień jak moralność, seksualność czy emocje, jest z jednej strony odpowiedzią na postępujący rozwój w dziedzinie algorytmów, ze szczególnym uwzględnieniem algorytmów sztucznej inteligencji, a z drugiej strony stwarza nowe obszary poddawane analizie i refleksji. Jednak ze względu na znaczące braki w wielkich zbiorach danych praca na nich jest obciążona ryzykiem błędu. Choć uczymy maszyny rozpoznawać, co to znaczy mieć więcej pieniędzy, to czy rzeczywiście uczymy je, czym jest bogactwo? Czy potrafimy wskazać i zmierzyć obszary biedy? Jak zoperacjonalizować wykluczenie, piękno czy niebezpieczeństwo? Jakiej moralności będziemy uczyć cyfrowych asystentów? Jeśli nie rozumiemy wszystkich ograniczeń, pełnego spektrum intencji gromadzenia, pozyskiwania i przetwarzania danych, których dostarczamy algorytmom, każdy aspekt naszej działalności powinien być poddawany regularnej i dogłębnej kontroli. Regresja, wnioskowanie statystyczne, korelacja, poszukiwanie wzorców – wszystko to, z pewnym prawdopodobieństwem, przy określonym przedziale ufności, rozpościera swoją moc przewidywania na znakomitą większość wydarzeń, sytuacji, obrotów spraw, fenomenów. Szczególnie jednak interesujące jest to, gdzie aktualnie leży granica mocy predykcyjnej stworzonych modeli. Zyskiem ze zwiększania rozdzielczości danych, które o sobie zbieramy, może być lepsze zrozumienie nas samych (niezależnie od tego, czy robi to sprytny analityk, czy krótki kod uczenia maszynowego). Stratne są zaś osoby nieprzystające do modelu obrzeża społecznego i kulturowego *mainstreamu*, systemowo lub chwilowo wykluczeni odmienicy. Dopóki nie zbudujemy bazy danych kompletnych, widocznych i widzialnych, transparentnych – według pomysłu Borgego oplatających nasze życia w skali 1:1 – ciągle będzie nam towarzyszyć widmo kosztów rozwiązań algorytmicznych.

---

<sup>72</sup> Por. C. O'Neil, *Weapons...*, dz. cyt.

<sup>73</sup> K. Crawford, *The hidden biases in Big Data*, „Harvard Business Review”, 1 kwietnia 2013, <http://hbr.org/2013/04/the-hidden-biases-in-big-data> (8 lutego 2019).

## BIBLIOGRAFIA

- Assmann, Aleida. *Między historią a pamięcią. Antologia*. Tłum. Zofia Dziewanowska-Stefańczyk i in. Warszawa: Wydawnictwa UW, 2013.
- Awad, Edmond. *Moral Machine. Perception of Moral Judgment Made by Machines*. Massachusetts Institute of Technology. Czerwiec 2017. <https://dam-prod.media.mit.edu/x/2017/06/18/awad-ms-17.pdf>.
- boyd, danah, Kate Crawford. „Six provocations for Big Data”. *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*. Wrzesień 2011. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1926431](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431).
- Buolamwini, Joy, Timnit Gebru. „Gender shades. Intersectional accuracy disparities in commercial gender classification”. *Proceedings of Machine Learning Research* 81 (2018).
- Cardon, Dominique. „Deconstructing the algorithm. Four types of digital information calculations”. W: *Algorithmic Cultures. Essays on Meaning, Performance, and New Technologies*, red. Robert Seyfert, Jonathan Roberge. London–New York: Routledge, 2016.
- Crawford, Kate, Vladan Joler. „Anatomy of an AI system”. 2018. [www.anatomyof.ai](http://www.anatomyof.ai).
- Danks, David, Alex J. London. „Algorithmic bias in autonomous systems”. W: *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI 2017)*. 2017.
- Eubanks, Virginia. *Automating Inequality. How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin’s Press, 2018.
- Gillespie, Tarleton. *Algorithm (Digital Keywords)*. Culture Digitally. 25 czerwca 2014. <http://culturedigitally.org/2014/06/algorithm-draft-digitalkeyword/>.
- Gitelman, Lisa, Virginia Jackson. *Introduction*. W: *„Raw Data” Is an Oxymoron*, red. Lisa Gitelman. Cambridge–London: MIT Press, 2013.
- Greenfield, Adam. *Radical Technologies. The Design of Everyday Life*. London–New York: Verso, 2017.
- Magalhães, João. *Algorithmic visibility. Elements of new regime of visibility*. Materiały konferencyjne. Kwiecień 2017. [https://www.researchgate.net/publication/321245421\\_Algorithmic\\_visibility\\_-\\_Elements\\_of\\_new\\_regime\\_of\\_visibility](https://www.researchgate.net/publication/321245421_Algorithmic_visibility_-_Elements_of_new_regime_of_visibility).
- Mittelstadt, Brent D., Patrick Allo, Mariarosaria Taddeo, Sandra Wahter, Luciano Floridi. „The ethics of algorithms. Mapping the debate”. *Big Data & Society*, 1 grudnia 2016.
- Sandvig, Christian. *Seeing the sort. The aesthetic and industrial defense of ‘the algorithm’*. „Media-N”. Styczeń 2014. <http://median.newmediacaucus.org/art-infrastructures-information/seeing-the-sort-the-aesthetic-and-industrial-defense-of-the-algorithm/>.
- Wang, Yilun, Michał Kosiński. *Deep neural networks are more accurate than humans at detecting sexual orientation from facial images*. PsyArXiv. 7 września 2017. <https://psyarxiv.com/hv28a/>.

Data wpłynięcia: 7 grudnia 2018 r. Data zatwierdzenia do druku: 26 marca 2019 r.

## WHAT DO ALGORITHMS SEE? CONSEQUENCES OF ALGORITHMIC (IN)VISIBILITY AND (IN)TRANSPARENCY OF DATA

Algorithms and their performance are a derivative of the efficiency of their design, quality of the data they use as well as the data that they tend to omit – intentionally or due to lack of access to it. Equally important are assumptions made by algorithms to fill the gaps in the data on which they work. As technological and social constructs they are reality-describing models which either remain visible in the process of data analysis or are pushed beyond the limits of our perception. The author discusses ways in which data can be made both visible and transparent, while diagnosing the sources of data omissions and their consequences.

**SŁOWA KLUCZOWE:** algorytm, dane, widoczność/widzialność danych, *proxies*

**KEY WORDS:** algorithm, data, data visibility/transparency, *proxies*

