

---

# Dociekania

---

## Rządy ludzkie czy algorytmiczne? O automatyzacji władzy sądenia

---

Ewa Płonowska-Ziarek

---

TEKSTY DRUGIE 2021, NR 6, S. 237–252

---

DOI: 10.18318/td.2021.6.15

---

**W** okresie pandemii COVID-19, rosnącej ksenofobii oraz homo- i transfobii łatwo przeoczyć mniej nagłośnioną, aczkolwiek równie niebezpieczną globalną pandemię spowodowaną przez postępującą „datafikację” oraz idące z nią w parze przekształcenia obliczeniowe codziennego życia, możliwe, że nieodwracalne<sup>1</sup>. Jako że coraz więcej codziennych aktywności człowieka – począwszy od randek, prowadzenia samochodu, rozrywki i wyszukiwania w internecie, aż po wybór i klasyfikowanie informacji, pracę, bankowość, opiekę zdrowotną i relacje społeczne – jest mediowanych przez algorytmy, a nie ludzkie decyzje i debaty, trudno przewidzieć dalekosiężne konsekwencje owych zmian, zwłaszcza w sferze demokracji oraz sprawiedliwości płciowej, etnicznej i ekonomicznej.

Najbardziej oczywistą trudność stanowi niedostępność spowodowana tajnością zastrzeżonych algorytmów

---

**Ewa Płonowska-Ziarek**

– prof. na wydziale Comparative Literature, the University at Buffalo. Autorka wielu artykułów i książek o feministycznej teorii politycznej, teorii rasy, modernizmie, między innymi: *Arendt, Natality and Biopolitics* (wraz z Rosalyn Diprose, 2019), *Feminist Aesthetics and the Politics of Modernism* (2012); *An Ethics of Dissensus: Feminism, Postmodernity, and the Politics of Radical Democracy* (2001); *The Rhetoric of Failure: Deconstruction of Skepticism, Reinvention of Modernism* (1995).

---

1 E. Ziarek *Triple Pandemics: COVID-19, Anti-Black Violence, and Digital Capitalism*, „Philosophy Today” 2020 vol. 64 no. 4, s. 925-930.

oraz brakiem specjalistycznej wiedzy technicznej wśród społeczeństwa, które jest coraz bardziej uzależnione od algorytmicznych decyzji. Napędzane algorytmicznie globalne praktyki gromadzenia danych, profilowania użytkowników, inwigilacji oraz analiz predykcyjnych stosowane przez gigantów technologii cyfrowych – Amazon, Google, Facebook, Microsoft, Apple – nie podlegają publicznej kontroli i dlatego są zaliczane do „black box”<sup>2</sup>. Ta globalna hegemonia prywatnych korporacji cyfrowych nie tylko rodzi pytania o ochronę danych i wzmoczoną inwigilację, ale także grozi przekształceniem polityki i walki o sprawiedliwość w to, co uczeni i filozofowie nazwali „algorytmiczną rządomyślnością”<sup>3</sup>, a więc w porządek charakteryzujący się coraz częstszym zastępowaniem decyzji politycznych i prawnych przez zautomatyzowane algorytmy. W przeciwieństwie do politycznych procesów deliberacyjnych decyzje algorytmiczne klasyfikują jednostki i dystrybuują dobra publiczne na podstawie profilowania cyfrowego. Kontynuując i automatyzując opresję ze względu na kolor skóry, płeć i status ekonomiczny, profilowanie obliczeniowe stanowi w istocie cyfrowe unowocześnienie długoletniej politycznej praktyki inwigilowania mniejszości etnicznych i rasowych oraz działaczy dysydenckich.

Przyspieszona przez pandemię COVID-19 algorytmiczna rządomyślność idzie w parze z cyfrowym kapitalizmem. Na przykład znaczne „uzdalnienie” pracy i edukacji przyniosło nowe podziały oraz dalszą prekaryzację. Telepraca stoi w sprzeczności z zawodami, których nie da się przenieść do sieci i które są albo coraz częściej likwidowane, albo wykonywane przez „niezbędnych pracowników” w niebezpiecznych warunkach, bez odpowiedniego sprzętu ochronnego. Nigdy wcześniej przymiotnik „niezbędny” nie był tak blisko powiązany z „prekaryjnością”. Dodajmy do tego agresywny nacisk na dalszą automatyzację w celu zastąpienia utraconych miejsc pracy uczeniem maszynowym<sup>4</sup>. Te wstrząsy ekonomiczne w obszarze zatrudnienia przyspieszyły trwającą już mutację neoliberalizmu w kapitalizm komputacyjny (Stiegler) lub kapitalizm inwigilacji (Zuboff). Według Zuboff ten rodzaj kapitału jest napędzany imperatywem wzmoczonej akumulacji danych gromadzonych w wyniku śledzenia aktywności

2 F. Pasquale *The Black Box Society*, Harvard University Press, Cambridge 2015.

3 B. Stiegler *The Age of Disruption. Technology and Madness in Computational Capitalism*, trans. D. Ross, Polity Press, Cambridge 2019; A. Rouvroy, T. Berns, *Algorithmic Governmentality and Prospects of Emancipation*, „Réseaux” 2013 vol. 177 no. 1, s. 163-196.

4 S. Lynch *How Work Will Change Following the Pandemic*, <https://hai.stanford.edu/blog/how-work-will-change-following-pandemic> (dostęp: 16 lipca 2020).

użytkowników w sieci. Dane są następnie wykorzystywane do wytwarzania produktów prognostycznych, począwszy od reklam po predykcyjne narzędzia policyjne, a następnie spieniężane. Mechanizm kapitalizmu inwigilacji, stworzony przez Google i obecnie udoskonalany przez wszystkich cyfrowych gigantów działających na skalę globalną, polega na wydobywaniu coraz większej ilości danych „pochodnych”, pozostawianych przez użytkowników w sieci jako produkt uboczny miliardów transakcji online oraz interakcji z mediami społecznościowymi, urządzeniami typu „smart” i wyszukiwarkami internetowymi – są to na przykład informacje o słownictwie i pisowni fraz używanych do wyszukiwania, częstości i czasie ich wpisywania, odwiedzanych stronach, liczbie kliknięć w reklamy itd. Część tych danych reinwestuje się w celu ulepszenia usług internetowych, ale większość stanowi lukratywną „nadwyżkę behawioralną” wykorzystywaną nie tylko do dopasowywania reklam do informacji o profilu użytkownika (UPI), ale przede wszystkim do tworzenia nowych, dokładniejszych produktów prognostycznych, które przewidują nasze działania i z kolei stają się przedmiotem obrotu na nowym rynku prognoz behawioralnych<sup>5</sup>. Jak wskazuje coraz więcej badaczy – m.in. Safiya Umoja Noble, Simone Brown czy Bonnie Sheehey – intensywność cyfrowego nadzoru i sposób profilowania wraz z ich konsekwencjami ekonomicznymi znacząco różnią się od siebie w zależności od lokalizacji geopolitycznej, a także koloru skóry, pochodzenia etnicznego, płci i poziomu ubóstwa.

W kontekście współczesnych zagrożeń dla demokracji, jakie stanowią nowe technologie władzy wyłaniające się z połączenia (lub skrzyżowania) cyfrowego kapitalizmu „big data” i sztucznej inteligencji, chciałabym poruszyć problem automatyzacji władzy sądenia. Zagrożenia te są przedmiotem dyskusji interdyscyplinarnych inicjowanych między innymi przez O’Neil<sup>6</sup>, Noble<sup>7</sup>, Zuboff<sup>8</sup>, Benjamin<sup>9</sup>, Kuehn i Saltera<sup>10</sup> (2020). Podtytuł książki O’Neil

---

5 S. Zuboff *Wiek kapitalizmu inwigilacji*, przeł. A. Untershuetz, Zysk i S-ka, Poznań 2020.

6 C. O’Neil *Weapons of Math Destruction*, Crown Publishing, New York 2016.

7 S. Noble *Algorithms of Oppression: How Search Engines Reinforce Racism*, New York University Press, New York 2018.

8 S. Zuboff *Wiek kapitalizmu inwigilacji*.

9 R. Benjamin *Race After Technology: Abolitionist Tools for the New Jim Code*, Polity Press, Cambridge 2019.

10 K.M. Kuehn, L.A. Salter *Assessing Digital Threats to Democracy, and Workable Solutions: A Review of the Recent Literature*, „International Journal of Communication” 2020 no. 14, s. 2589–2610.

*Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Broń matematycznej zagłady: Jak algorytmy zwiększają nierówność i zagrażają demokracji) w skrócie oddaje główny zakres tych debat. To, co chcę dodać do tej dyskusji, to namysł nad polityczną i epistemiczną słusznością przekształcania idei jakościowych, takich jak uczciwość (*fairness*), sprawiedliwość, skuteczność nauczania, redukcja przestępczości czy szczęście, w wymierne, policzalne wielkości lub dane. Właśnie takie przekształcanie umożliwia zastąpienie sądów procedurami algorytmicznymi.

### **Część I: Rządy ludzi czy algorytmów?**

W książce *Undoing the Demos* wydanej w roku 2015 Wendy Brown zwięźle podsumowuje podstawowe założenie demokracji jako „formy politycznej, w której ogół społeczeństwa zarządza wspólnotą polityczną, a tym samym sobą”<sup>11</sup>. Ta forma rządzenia była interpretowana i historycznie wdrażana na różne sposoby, miała postać demokracji liberalnej, kapitalistycznej, socjalistycznej, bezpośredniej, reprezentatywnej czy deliberatywnej. Brown słusznie wskazuje, że pomimo tych różnic demokracja wymaga co najmniej, aby „ludzie autoryzowali własne prawa i główne decyzje polityczne, czy to bezpośrednio, czy za pośrednictwem wybranych przedstawicieli”, oraz aby posiadali pewne środki i możliwości uczestnictwa w głównych instytucjach – publicznych, prawnych, kulturalnych czy ekonomicznych – które zarządzają ich życiem.

Nawet w najbardziej sprzyjających okolicznościach ideał demokratycznych rządów i decyzji splata się z ciągłą debatą publiczną i sprzeczną interpretacją wartości oraz praw. Jednak ze względu na sprzeczności wynikające ze współistnienia z jednej strony ideałów demokratycznej samorządności, wolności i równości, a z drugiej – realnych historycznych praktyk ucisku, wykluczenia i opresji (począwszy od nierówności ekonomicznych oraz dyskryminacji ze względu na płeć, kolor skóry czy religię, na przykład islamofobii, po ksenofobię) wiele teoretyczek i teoretyków demokracji spod znaku feminizmu, krytycznej teorii „rasowości” (*critical race theory*) i politologii zwraca uwagę, że polityka demokracji jest nierozzerwalnie związana z aktywizmem, walką z wykluczeniem i dyskryminacją oraz kwestionowaniem dyskryminujących praw i ich wdrażania.

Mówiąc pośrednio i bezpośrednio o tych nierównościach, Brown wymienia również wstępne warunki istnienia demokracji, które jako minimum

11 W. Brown *Undoing the Demos*, Zone Books, New York 2015, s. 178.

powinny obejmować redystrybucję środków materialnych w celu zmniejszenia nierówności ekonomicznych, zabezpieczenie możliwie jak najszerszej partycypacji politycznej oraz wspieranie edukacji publicznej nastawionej na obywatelskość i dobro publiczne. Taka edukacja powinna zwiększać wiedzę na temat tego, w jaki sposób „władza, historia, reprezentacja i sprawiedliwość” funkcjonują w życiu publicznym<sup>12</sup>. Wszystkie elementy tego bardzo krótkiego zarysu demokracji – samorządność, wzgląd na dobro publiczne, debaty na temat sprawiedliwości oraz, podkreśliłabym, kontestacja niesprawiedliwości i dyskryminacji – wyraźnie akcentują znaczenie deliberacji, osądów i zaangażowania w życie polityczne. Zdaniem Brown politykę demokracji niweczą neoliberalna deregulacja rynków i niszczenie instytucji publicznych tworzących ramy dla redystrybucji ekonomicznej, edukacji publicznej i zaangażowania politycznego. Jednakże w roku 2021 neoliberalne zagrożenie dla demokracji spotęgowały nie tylko rosnący populizm, kryzys instytucji demokratycznych i autorytaryzm, ale także kapitalizm sieci oraz technologie cyfrowe, które wpływają na wszystkie aspekty naszego codziennego życia. Jak jednak słusznie zauważają Lucy Bernholz, Hélène Landemore i Rob Reich, kwestia „głębokiego wpływu”, jaki stale aktualizowane technologie cyfrowe wywierają na instytucje demokratyczne, prawa człowieka i obywatelskość, nie została jeszcze wystarczająco skonfrontowana przez teoretyków demokracji, którzy mają skłonność do skupiania się na nietechnologicznych lub niskotechnologicznych (*low-tech*) przyczynach powszechnego kryzysu demokracji<sup>13</sup>. Jak będę dowodzić, ta niebezpieczna synergia między cyfrowym kapitalizmem a cyfrowymi technologiami władzy prowadzi do nowej hybrydowej formacji politycznej, a mianowicie rządzenia ludźmi według algorytmów.

Jedną z najbardziej charakterystycznych cech tej hybrydowej formacji stanowi „outsourcing” czy przekazanie decyzji i sądów politycznych systemom algorytmicznym.

Chciałabym, żeby fraza „władza algorytmu” (*government by algorithm*) była albo moim pomysłem, albo przyciągającym uwagę sensacyjnym nagłówkiem, ale niestety tak nie jest. Nawiązując do jednego z najsłynniejszych haseł w historii amerykańskiej polityki: „władza ludu, sprawowana przez ludzi i dla ludzi”, fraza „władza algorytmu” to w rzeczywistości część tytułu raportu zleconego przez Konferencję Administracyjną Stanów Zjednoczonych

<sup>12</sup> Tamże, s. 179.

<sup>13</sup> *Introduction*, w: *Digital Technology and Democratic Theory*, ed. L. Bernholz, H. Landemore, R. Reich, The University of Chicago Press, Chicago 2021, s. 3.

i przedstawionego w lutym 2020 roku przez dużą interdyscyplinarną grupę prawników, informatyków i socjologów z Uniwersytetu Stanforda i Uniwersytetu Nowojorskiego. Tytuł raportu brzmi *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies* (Władza algorytmu: sztuczna inteligencja w federalnych agencjach administracyjnych). Jest wiele aspektów tego raportu, które niepokoją: po pierwsze, ma on zaradzić ignorancji i dezorientacji związanej z chaotycznym, nieuporządkowanym wdrażaniem algorytmów (czyli technologii zarządzania wynalezionych przez prywatne korporacje technologiczne) na wielu szczeblach władzy. Twierdzenie, że „niewiele wiemy o tym, jak agencje rządowe wykorzystują sztuczną inteligencję”, jest szczególnie niepokojące w obliczu publicznych apeli o zwiększenie regulacji prawnych w zakresie wykorzystania sztucznej inteligencji przez korporacyjnych gigantów technologicznych. Na ile możemy mieć zaufanie do regulacji prawnych dotyczących pozaprawnego działania sztucznej inteligencji i uczenia maszynowego, jeśli sam rząd nie wie, co robi, a zatem nie może sprawdzić, czy jego działania są zgodne z zasadami i prawami instytucjonalnymi, takimi jak prawo do rzetelnego procesu prawnego (*due process*) czy ochrony danych osobowych? Po drugie, raport niejako potwierdza brak przejrzystości w instytucjach publicznych oraz praktyki tzw. black boxing i dlatego jego autorzy wskazują, że ich ustalenia ogranicza to, że „opierali się na publicznie dostępnych źródłach”, które „rzadko dostarczają wystarczająco szczegółowych informacji technicznych na temat systemów sztucznej inteligencji i oferują niewielki wgląd” w procesy gromadzenia informacji w takich przypadkach<sup>14</sup>. Jeśli wiedza ekspertów technicznych i prawnych na temat roli, jaką decyzje algorytmiczne odgrywają w instytucjach publicznych, jest ograniczona, to co mają powiedzieć obywatele tacy jak ja, którzy nie są ekspertami? Brakuje nie tylko spójnej polityki dotyczącej wykorzystania sztucznej inteligencji i uczenia maszynowego w polityce i życiu publicznym, ale również wiedzy na temat tego, jak władza wykorzystuje takie zautomatyzowane decyzje. W związku z tym głównym celem raportu jest, że tak powiem, ustalenie po fakcie, w jak szerokim zakresie agencje rządowe wdrażają rozwiązania oparte na sztucznej inteligencji i uczeniu maszynowym oraz jak te narzędzia zostały pozyskane (to znaczy, czy zostały opracowane przez daną instytucję, czy

14 D.F. Engstrom, D.E. Ho, C.M. Sharkey, M.-F. Cuéllar *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies*, Report Submitted to the Administrative Conference of the United States, luty 2020, <https://www-cdn.law.stanford.edu/wp-content/uploads/2020/02/ACUS-AI-Report.pdf>, s. 12–13.

przez prywatnych kontrahentów). Po zbadaniu sytuacji w stu czterdziestu dwóch najważniejszych departamentach federalnych w raporcie stwierdzono, że prawie połowa z nich z wykorzystuje sztuczną inteligencję w obszarach obejmujących zdrowie, egzekucję prawa, regulacje finansowe oraz – co jest być może najbardziej niepokojące – w obszarze komunikacji ze społeczeństwem „dotyczącej jego praw i obowiązków”<sup>15</sup>. Decyzje podejmowane przez algorytmy są szczególnie widoczne w odniesieniu do egzekwowania prawa (zwłaszcza w sprawach karnych), imigracji i patrolowania granic, a także podziału i alokacji zasobów materialnych i ekonomicznych (czyli sprawiedliwości dystrybucyjnej).

Chociaż geneza owych praktyk sięga lat 90., w raporcie skupiono się na najbardziej oczywistych, nowych, i niepokojących aspektach algorytmicznego zarządzania. Pierwszym jest brak zrozumienia złożoności tych narzędzi (na przykład zaawansowanych sieci neuronowych) nawet przez tych, którzy je wynaleźli<sup>16</sup>. Drugim fakt, że zautomatyzowane decyzje „funkcjonują zgodnie z regułami, które są tak złożone”, że nie tylko wymykają się praktycznej kontroli, ale, co bardziej alarmujące, „nie zgadzają się z żadnymi realnymi ludzkimi przekonaniem dotyczącymi tego, jak działa świat, lub po prostu wykraczają poza rozumowanie na ludzką skalę”<sup>17</sup>. Ponieważ decyzje algorytmiczne zakłócają ludzkie rozumienie świata i proces nadawania mu sensu, uważam, że zarządzanie algorytmiczne stoi w fundamentalnej sprzeczności z demokratyczną zasadą uczestnictwa w zbiorowej samorządności.

Jak między wierszami przyznają sami autorzy raportu, decyzje algorytmiczne podważają podstawową zasadę demokracji deliberatywnej, którą stanowią polityczne praktyki uzasadniania postanowień i wyjaśniania kryteriów sądenia: „rosnące skomplikowanie i moc sztucznej inteligencji popychają agencje (federalne) w kierunku w pełni zautomatyzowanego procesu podejmowania decyzji, pozostawiając stopniowo coraz mniej miejsca dla ludzkiej rozważgi i osądu”<sup>18</sup>. Autorzy raportu w pełni zdają sobie sprawę, że automatyzacja procesu sądenia i podejmowania decyzji nie tylko rodzi nowe pytania o sprawiedliwość, przejrzystość i odpowiedzialność za dyskryminację<sup>19</sup>, ale

---

15 Tamże, s. 6.

16 Tamże, s. 11.

17 Tamże.

18 Tamże.

19 Tamże, s. 13.

również podważa praktykę „podawania powodów”, to znaczy uzasadniania i wyjaśniania kryteriów osądów i decyzji politycznych. Praktyka ta stanowi zagrożenie dla praw człowieka, sprawiedliwości i rzetelnego procesu (*due process*). Podstawowym pytaniem, którego autorzy raportu nie stawiają wprost, jest to, czy władza algorytmiczna jest nadal władzą demokratyczną, czyli czy jest ona zgodna z zasadami demokracji partycypacyjnej rozumianej jako rządu ludu. A jeśli nie, to jaka forma polityczna wyłania się z rządzenia ludźmi według algorytmu? „Społeczeństwo black box”? „Algorytmiczna rządomyślność”? „Kapitalizm inwigilacyjny”? „Algorytmiczny Lewiatan?” A może coś zupełnie innego?

## **Część II. O zasadności zastępowania władzy sądenia przez algorytmy**

Teoria demokracji traktuje osądy polityczne, bez względu na to, jak są one rozumiane i definiowane, jako nieodzowny element debaty publicznej i aktywizmu. Dlatego właśnie „zdolność dokonywania krytycznych i refleksyjnych osądów stanowi główną cechę nowoczesnego obywatelstwa demokratycznego”, by użyć trafnego sformułowania Lindy Zerilli<sup>20</sup>. Ponieważ polityka demokratyczna jest nie do pomyślenia bez formułowania i uzasadniania sądów, gdy teoria polityczna (jak wskazuje Zerilli w artykule z 2018 roku zatytułowanym *Toward a Democratic Theory of Judgment*) zajmuje się władzą sądenia jako taką, skupia się przede wszystkim na: 1) kryteriach solidnych, rozsądnych sądów (na przykład czy powinny one ograniczać się jedynie do racjonalności, czy zostać poszerzone o afekt lub wartości kulturowe), 2) trudnościach dokonywania osądów w społeczeństwach wielokulturowych w przypadku braku wspólnych norm, 3) trudnościach procesu sądenia w kontekście konfliktów politycznych itd.<sup>21</sup> Jeśli teoria polityczna w ogóle rozważa rezygnację z sądów, to głównie w kontekście zagrożeń dla demokracji, począwszy od redukcji polityki do wydajności aparatu administracyjnego, po autorytaryzm i banalność zła (Arendt).

Dlaczego władza sądenia jest tak ważna? Niezależnie od tego, czy mówimy o sądach epistemicznych, moralnych czy politycznych, pozwalają nam one łączyć ogólne prawa, koncepcje i normy z konkretnymi przypadkami, które mogą im odpowiadać, i odwrotnie. Nawet w najbardziej racjonalnym

20 L. Zerilli *Toward a Democratic Theory of Judgment*, w: *Judgment and Action*, ed. V. Soni, T. Pfau, Northwestern University Press, Evanston 2018, s. 191.

21 Tamże.



ujęciu władzy sądenia, zaproponowanym na przykład przez XVIII-wiecznego filozofa Immanuela Kanta, zwyczajne sądy – to jest dobre, to jest kot – nie są wcale prostymi procedurami poznawczymi, które stosuje się mechanicznie. W rzeczywistości jest to rodzaj sztuki, która pozwala łączyć postrzeganie zmysłowe z racjonalnymi pojęciami dzięki pośrednictwu wyobraźni. W konsekwencji nawet racjonalne sądy różnią się od wiedzy teoretycznej lub empirycznej. Tę różnicę między wiedzą a sądzeniem widać najlepiej w przypadku sądów refleksyjnych i estetycznych, opartych raczej na przyjemności i niezadowoleniu (lub nieprzyjemności) niż na pojęciach. Ale, jak twierdzi Arendt, jest ona również kluczowa do zrozumienia wydarzeń historycznych i politycznych, które wykraczają poza dotychczasowe pojęcia i kategorie myślenia lub je niszczą. Sądy, mimo że nie bazują na pojęciach, są nadal komunikowalne, ponieważ odwołują się do perspektywy innych, co Kant nazywa *sensus communis*, co dosłownie tłumaczy się jako „zmysł wspólny”, a Arendt – mentalnością rozszerzoną. Z powodu tej różnicy między władzą sądenia a wiedzą Kant (w *Krytyce czystego rozumu*) zwraca uwagę, że nawet najlepiej wykształceni ludzie mogą nie posiadać zdolności sądenia. Dlatego lekarze, prawnicy, sędziowie, naukowcy czy filozofowie, mimo że są ekspertami w swoich dziedzinach, mogą nie być w stanie odpowiednio zastosować sądów w określonych sytuacjach. I na odwrót, ludzie bez formalnego wykształcenia mogą przodować w wydawaniu sądów. Ten niedostatek u wykształconych elit Kant nazywa głupotą i zauważa, że nie jest niczym niezwykłym spotkać uczonych, którzy stosując wiedzę naukową, ponoszą fiasko w obszarze sądenia.

Reinterpretując Kanta, Arendt analizuje nową formę głupoty charakterystyczną dla współczesności<sup>22</sup>. Jak pisze, głównym jej przejawem jest w naszej epoce utrata władzy sądenia oraz *sensus communis* i zastąpienie ich logicznością<sup>23</sup>. Arendt używa tu określenia *logicality* (logiczność), aby odróżnić to zjawisko od uzasadnionych zastosowań logiki. Zastąpienie sądu logicznością jest możliwe, ponieważ logika również jest umiejętnością wspólną; nie jest ona jednak zdolna do rozumienia historycznego i do komunikowania znaczenia<sup>24</sup>. Logiczność zakłada bowiem spójną formalną prawdę, którą formułuje się w oderwaniu od doświadczenia, historii, nieprzewidywalnych wydarzeń oraz relacji z innymi. Taka oderwana od rzeczywistości logiczność staje się

22 H. Arendt *Rozumienie a polityka*, przeł. J. Sieradzki, „Literatura na Świecie” 1985 nr 6 (167).

23 Tamże.

24 Tamże.

głupotą, kiedy stosuje się ją jako hegemoniczne kryterium rozumienia historycznego i politycznego. Jak się przekonamy, logiczna głupota leży u podstaw zastępowania ludzkich sądów automatycznymi decyzjami algorytmicznymi. Być może moglibyśmy nazwać ten proces „algorytmiczną głupotą”.

Skoro nawet proste sądy funkcjonujące w ramach przyjętych kryteriów naukowych, prawnych i normatywnych są często skazane na niepowodzenie, o wiele trudniej jest je formułować, gdy te kryteria są kwestionowane. A przecież to na tym właśnie polega polityczna i etyczna rola krytyki, która w odróżnieniu od logiczności zakłada, że władza sądenia nie wykracza poza kontekst społeczno-polityczny, wręcz przeciwnie, jest uwikłana w sieci relacji władzy i wiedzy. W związku z tym musimy ciągle mierzyć się z koniecznością odróżniania naszych sądów od uprzedzeń, często udających naukę, a także z trudnościami, jakich to nastrocza, zwłaszcza jeśli te społeczno-polityczne uprzedzenia wynikają z nierówności strukturalnych i dominacji. Nie wierzę, że możemy sami te uprzedzenia pokonać – to dlatego krytyczna praktyka analizowania sądów i uprzedzeń jest zawsze osadzona w debacie publicznej, kontestacji i wiedzy historycznej. Pokonywanie uprzedzeń jest uzależnione od trudnej lekcji, jaką zapewniają nam inni, kiedy podważają hegemoniczne wartości, a także nasze opinie i decyzje. Jak wskazuje Arendt, publiczny charakter sądów, łącznie z ich komunikowaniem i objaśnianiem innym, którzy mogą je odrzucić, stanowi kluczową cechę myśli politycznej, praktyki krytycznej oraz silnej sfery publicznej. Według Diprose i Ziarek<sup>25</sup> tak naprawdę wszystkie sądy polityczne są zwrócone ku innym – nie tylko dlatego, że trzeba je ciągle uzasadniać, ale przede wszystkim dlatego, że opinie innych, faktyczne lub potencjalne, bierzemy pod uwagę także przy formułowaniu i uzasadnianiu własnych sądów. Zgadza się z interpretacją Zerilli, według której w polityce demokratycznej „proces sądenia może się rozpocząć, dopiero gdy weźmie się pod uwagę punkt widzenia innych”<sup>26</sup>. Idąc za Arendt, twierdzą, że takie rozumienie władzy sądenia rzeczywiście stanowi jedną z najwyższych wartości politycznych w demokracji, ponieważ zakłada, że dzielimy świat z innymi, do których zwracamy się z naszymi osądami. Jak uczy nas historia, takie rozumienie wspólnego świata może zostać zniszczone nie tylko przez wykluczenie, zagładę, kolonializm osadniczy, rasizm i homofobię, ale również przez „logiczność”, która prowadzi do automatyzacji tych form opresji.

25 R. Diprose, E. Plonowska-Ziarek *Arendt, Natality and Biopolitics: Toward Democratic Plurality and Reproductive Justice*, Edinburgh University Press, Edinburgh 2018, s. 299-304.

26 L. Zerilli *Toward a Democratic Theory*, s. 207.

Jeśli uzasadnianie i kwestionowanie sądów stanowi kluczowy element demokratycznych rządów i jeśli każda praktyka krytyczna nastawiona na odróżnienie sądów od uprzedzeń zależy od uwzględnienia odpowiedzi i sądów innych, to „ideologia informatyzmu” (Golumbia) pozwala obejść te trudności przez powierzenie sądów sztucznej inteligencji. Po co poświęcać czas i wysiłek na doskonalenie i uzasadnianie naszych sądów, skoro przekazanie tej kompetencji procedurom algorytmicznym jest wydajniejsze i tańsze? Czyż narzędzia sztucznej inteligencji, takie jak niesławny COMPASS, opracowany w 1998 roku w Stanach Zjednoczonych przez firmę Northpointe w celu oceny ryzyka recydywy, nie zostały wymyślone po to, żeby wykształcone osoby – sędziów, lekarzy, prawników, a także nauczycieli, menedżerów do spraw rekrutacji, pracowników uniwersytetów prowadzących nabór na studia, urzędników państwowych czy badaczy literatury – uwolnić od trudności rozumienia i odpowiedzialności związanej z sądem?

Skoro teoria polityczna nie potrafi sobie wyobrazić w społeczeństwie demokratycznym możliwości zrzeczenia się odpowiedzialności za dokonywanie osądów, to czy badania informatyczne mogą nam pomóc w konfrontacji z konsekwencjami przekazania procesu sądenia sztucznej inteligencji? W następstwie szeroko zakrojonych debat na temat szkód i stronniczości systemów obliczeniowych informatycy również borykają się z problemem zasadności przekazania tej kompetencji systemom cyfrowym, mimo że etyka nie jest dla branży technologicznej najwyższym priorytetem. Uzasadnienie wyjściowe, że matematyczna neutralność sztucznej inteligencji i uczenia maszynowego może przeciwdziałać społeczno-politycznym uprzedzeniom i nierównościami, samo w sobie jest problematyczne, bo nawet gdyby to było możliwe, powtarza ono hiperracjonalistyczne marzenie o przekroczeniu ludzkich konfliktów, pragnień, ograniczeń i zakorzenienia w świecie. Co więcej, zastąpienie decyzji sztuczną inteligencją oznacza zrzeczenie się odpowiedzialności za historię ucisku, ludobójstwa i nierówności. Jednak jak nazbyt wyraźnie pokazują przełomowe prace O’Neil, Noble, Benjamin, Eubanks, Pasquale i Rouvroy, sztuczna inteligencja nie tylko reprodukuje krzywdy i dyskryminację, ale także utrudnia ich zwalczanie. Zidentyfikowano wiele przyczyn takiego stanu rzeczy. Po pierwsze, technologie nigdy nie są neutralnymi narzędziami, lecz raczej ekonomicznymi i społeczno-politycznymi operacjami władzy. Po drugie, wskaźniki danych społecznych wykorzystywane do szkolenia maszyn kształtuje wieloletnia historia systemowych niesprawiedliwości związanych z kolorem skóry, statusem ekonomicznym i płcią (co Ruha Benjamin nazywa „podwójnym kodowaniem”). Po trzecie,

alarmujące są również brak przejrzystości i wzmożona inwigilacja, do której owe technologie są wykorzystywane. Po czwarte, „big data” i sztuczna inteligencja tworzą nowe hierarchie władzy/wiedzy pomiędzy tymi, którzy są w posiadaniu kapitału ekonomicznego, politycznego i intelektualnego potrzebnego do projektowania modeli algorytmicznych, a społecznościami podporządkowanymi nieuregulowanym prawnie decyzjom algorytmicznym. Za Arendt dodałabym do tej listy również logiczność jako nową wartość hegemoniczną zastępującą *sensus communis*.

W następstwie toczącej się debaty na temat korzyści i szkód związanych ze stosowaniem sztucznej inteligencji w relacjach społeczno-politycznych informatycy i naukowcy zajmujący się krytycznymi studiami nad danymi wykazali istnienie dużej skali sądów oraz decyzji zakodowanych na każdym etapie procesu uczenia maszynowego (*Machine Learning Pipeline*). Jak wskazuje O’Neil, „słabe punkty (*blind spots*) tego modelu odzwierciedlają priorytety sądenia jego twórców [...]. Pomimo swojej rzekomej bezstronności stanowią odbicie konkretnych celów i ideologii”. Jak podsumowuje badaczka, modele algorytmiczne „są opiniami wbudowanymi w matematykę”<sup>27</sup>. Bazując na interdyscyplinarnych badaniach nad potencjalnymi szkodami wywołwanymi przez sztuczną inteligencję, chcę się tutaj skupić na węższej, ale kluczowej kwestii, a mianowicie na roli, jaką sądy (często niezauważane) odgrywają w przekształcaniu pojęć jakościowych w wymierne, policzalne wielkości lub dane. Stosunkowo niedawno pod wpływem metodologii wyłaniających się z ilościowych nauk społecznych (a nie filozofii czy teorii politycznej) w świecie informatyki zaczęły się pojawiać rozważania na temat epistemicznych i etycznych trudności związanych z transformacją jakościowych cech życia ludzkiego w pomiary matematyczne, zwłaszcza w przypadku pytań otwartych, takich jak: co stanowi dobre życie? Jakie powinny być zasady demokracji? Czym jest szczęście, wolność lub jak nadać sens światu? Niemniej jednak wpływ ten, obok rozważań dotyczących dokładności takich pomiarów, umożliwił postawienie bardziej istotnych pytań o trafność i zasadność tego typu transformacji, na przykład w jakich okolicznościach przekształcenie czegoś, czego nie można zmierzyć, w mierzalne substytuty lub wskaźniki zastępcze jest trafne, właściwe, i uzasadnione? Jakie są tego konsekwencje?

Dlaczego kwestia zasadności (*legitimacy*) jest dla mnie tak ważna? Bo ponownie wprowadza na teren informatyki (by tak rzec, tylnymi drzwiami) kwestię sądów moralnych, politycznych i epistemicznych. Jak na przykład

27 C. O’Neil *Weapons of Math Destruction*, s. 21.

argumentują Jacobs i Wallach, wszelka ocena stronniczości, uprzedzeń lub szkody w obszarze uczenia maszynowego jest niewystarczająca, jeśli koncentruje się tylko na rzetelności pomiarów (tzw. rzetelności operacyjnej) i nie kwestionuje trafności przekształcania badań jakościowych w ilościowe<sup>28</sup>. Ponieważ problem trafności i rzetelności pomiaru notorycznie się w informatyce myli lub łączy, problem uczciwości (*fairness*) jako idei jakościowej o wielorakich i spornych znaczeniach rozpatruje się redukcyjnie jedynie w kategoriach błędu statystycznego. Na przykład większość amerykańskich dyskusji na temat roli, jaką w algorytmicznych wskaźnikach ryzyka recydywy dla czarnych i białych przestępców odgrywają rasistowskie uprzedzenia, koncentruje się na kwestii fałszywie dodatnich lub ujemnych wyników statystycznych (*false positives or false negatives*) i nie zajmuje się głębiej kontekstem białej supremacji, w którym takie narzędzia algorytmiczne są rozwijane i wykorzystywane.

Najogólniej rzecz biorąc, problem rzetelnego pomiaru sprowadza się do kwestii tego, czy podobne dane wejściowe dla modelu wygenerują zbliżone dane wyjściowe. Problem trafności i zasadności jest jednak o wiele bardziej złożony, ponieważ stawia nas przed pytaniem o konsekwencje przekształceń jakościowych idei, których nie da się bezpośrednio zmierzyć, takich jak uczciwość, skuteczność nauczania czy – najczęściej omawiane w literaturze informatycznej – ryzyko recydywy, na ilościowe, wymierne wskaźniki zastępcze. W jakich okolicznościach używanie takich ilościowych wskaźników jest właściwe i uzasadnione? Jak to ujęły Jacobs i Wallach, „systemy obliczeniowe często zawierają niedostrzegalne konstrukty teoretyczne, takie jak status socjoekonomiczny [...] konstruktów takich nie można zmierzyć bezpośrednio – zamiast tego trzeba je wywnioskować z pomiarów obserwowalnych i dostrzegalnych właściwości, które są z nimi powiązane”<sup>29</sup>. Należy zwrócić uwagę, że cytat ten problematycznie utożsamia postrzeganie świata z pomiarem oraz redukuje niemierzalne, ale powszechne idee – takie jak sprawiedliwość, władza polityczna, potrzeby medyczne – do abstrakcyjnych „konstruktów teoretycznych”, mimo że są w rzeczywistości częścią naszych codziennych doświadczeń i aspiracji. Pomimo tej błędnej interpretacji Jacobs i Wallach przyznają jednak, że proces przekształcania idei i wartości w mierzalne, ilościowe wskaźniki „z konieczności wiąże się z przyjmowaniem

28 A.Z. Jacobs, H. Wallach *Measurement and Fairness*, ACM Conference on Fairness, Accountability, and Transparency, marzec 2021, New York, US, <https://doi.org/10.1145/3442188.3445901>.

29 Tamże, par. 1.

założeń”, a zatem wprowadza „potencjalne rozbieżności między teoretycznym zrozumieniem konstrukcji, która ma być mierzona, a jej operacjonalizacją czy zastosowaniem w danym modelu pomiaru”<sup>30</sup>.

Chociaż rozróżnienie rzetelnego pomiaru i jego trafności jest ważne, to jednak nie kwestionuje ono w wystarczającym stopniu zasadności zastępowania niemierzalnych wartości i idei wartościami policzalnymi, na przykład używania dochodu jako wskaźnika zastępczego dla statusu społeczno-ekonomicznego, kosztów opieki jako wskaźnika potrzeb pacjenta, czy co gorsza, tak częstego w naszej dyscyplinie traktowania wyników testów jako wskaźnika sukcesu nauczania. Takie pomiary z zastosowaniem wskaźników zastępczych leżą wszak u podstaw zastępowania sądów przez procedury algorytmiczne. Algorytmizacja sądów ma daleko idące konsekwencje politycznych, takie jak stopniowe zastępowanie wartości moralnych, politycznych, historycznych i kulturowych punktacją lub rankingiem; za przykład może tu posłużyć wykorzystywanie punktowej oceny zdolności kredytowej jako wskaźnika wiarygodności danej osoby, traktowanie popularności w mediach społecznościowych jako odpowiednika dobra moralnego<sup>31</sup> czy zysku jako wskaźnika prawdy<sup>32</sup>.

Jeśli jednak wyjdziemy poza te ograniczenia wywodu Jacobs i Wallach, okaże się, że skupienie na zastępowaniu idei wartościami policzalnymi umożliwi bardziej zdecydowaną polityczną dyskusję na temat zasadności rządzenia ogółem ludzi za pomocą algorytmu i daleko idących konsekwencji takiej formy władzy. Rzeczywiście, jedno z najważniejszych pytań postawionych przez Jacobs i Wallach odnosi się do tzw. uzasadnienia konsekwencyjnego, czyli oceny społecznych, politycznych i moralnych skutków – zwłaszcza szkodliwych – zastępowania idei wartościami policzalnymi, a następnie pomiarami. Jak przyznają badaczki, „ocena uzasadnienia konsekwencyjnego często ujawnia szkody związane z niesprawiedliwością”<sup>33</sup>. I tu pojawia się najważniejszy, ale też najmniej rozwinięty punkt w ich dyskusji, w którym przyznają mimochodem, że pomiary nie tylko odzwierciedlają świat, w jakim żyjemy, ale w rzeczywistości również kształtują i zmieniają jego strukturę:

30 Tamże.

31 M. Broussard *Artificial Unintelligence: How Computers Misunderstand the World*, MIT Press, Cambridge, MA 2018, s. 11.

32 C. O’Neil *Weapons of Math Destruction*, s. 12.

33 A.Z. Jacobs, H. Wallach *Measurement and Fairness*, par. 3.2, 7.

ocena uzasadnienia konsekwencyjnego oznacza zatem odpowiedź na następujące pytania: Jak kształtuje się świat za pomocą pomiarów? W jakim świecie chcielibyśmy żyć? Jeśli istnieją sytuacje, w których konsekwencje stosowania pomiarów spowodowałyby naruszenie wartości, które chcielibyśmy utrzymać, wówczas w takich sytuacjach pomiarów nie należy stosować.<sup>34</sup>

Są to pytania poważne i głębokie, chociaż nie nowe, i trudno się dziwić, że niniejszy artykuł nie przynosi na nie odpowiedzi. Stawiają nas one w obliczu problemów politycznych i filozoficznych, których nie można rozwiązać za pomocą uczenia maszynowego. W jakim świecie chcemy żyć? Na co mamy nadzieję? Jak rozumiemy świat? Kim chcemy się stać? Jak rozumiemy wolność i sprawiedliwość? To pytania stanowiące podstawę kulturowych, politycznych, filozoficznych i religijnych refleksji na całym świecie od wieków. Tym, co się obecnie wyłania jako najbardziej naglący polityczny, etyczny i filozoficzny dylemat naszych czasów, jest pytanie o to, czy chcemy żyć w świecie rządonym przez algorytmy oraz czy chcemy żyć w świecie, w którym wydajność, przewidywalność i produktywność stają się substytutami sprawiedliwości społecznej. Parafrazując Arendt, możemy powiedzieć, że pytania te ujawniają nowe trudności w zrozumieniu algorytmicznej rządymyślności.

Przełożyła *Marta Aleksandrowicz*

---

34 Tamże.

## Abstract

---

**Ewa Płonowska-Ziarek**

UNIVERSITY AT BUFFALO

*Human or Algorithmic Governance? On the Automation of Judicial Power*

This article examines the issue of automation of the power of judging in the context of contemporary threats to democracy posed by the new technologies of power emerging from the combination of *big data* digital capitalism and artificial intelligence. The automation of judging and decision-making not only raises concerns about fairness, transparency and political accountability, but also confronts us with the fundamental question of whether algorithmic governance is still democratic governance, consistent with participatory democracy understood as the rule of the people. And if not, what political form emerges from governing people according to an algorithm?

## Keywords

---

threats to democracy, artificial intelligence, algorithmic governance, automation of judicial power, algorithmic decisions