

## INFORMACJE. PRZEGLĄDY. RECENZJE

Mirosław SZREDER

### O algorytmach Big Data

(na marginesie książki Cathy O’Neil pt. *Broń matematycznej zagłady. Jak algorytmy zwiększają nierówności i zagrażają demokracji*,  
Wydawnictwo Naukowe PWN, 2017)

Tempo, w jakim współcześnie zwiększają się możliwości gromadzenia i przetwarzania danych na temat wszystkiego, co robimy, o czym myślimy, z czym się utożsamiamy lub z czym się nie zgadzamy, ma prawo rodzić pytania o zagrożenia, które się z tym wiążą. Im więcej sfer życia społecznego i prywatnego objętych jest przetwarzaniem szczegółowych danych, tym większe rodzą się obawy. Jest to niepokój zarówno o to, kto i w jaki sposób wykorzystuje te informacje, jak i o to, czy zasilane nimi algorytmy nie zostaną wykorzystane przeciwko nam. I nie chodzi koniecznie o celowe posłużenie się nimi przeciwko komukolwiek (choć może do tego dojść), ale o konstrukcję takich modeli lub algorytmów, które sprzyjać będą jakiejś grupie osób, a dodatkowo mogą ten stan nierówności utrwaląć poprzez kolejne decyzje podejmowane na podstawie wskazań tych modeli. Jeszcze do niedawna, gdy moce obliczeniowe komputerów były znacznie słabsze i wypatrywaliśmy czasów, kiedy szybciej znajdziemy więcej informacji na interesujący nas temat, dostrzegaliśmy raczej szanse niż zagrożenia tej sytuacji. Obecnie skala obaw o powszechność stosowania algorytmów Big Data jest tak duża, że warto zadać pytania: może lepiej byłoby nie zbierać i nie przetwarzać wielu informacji o nas? Może lepiej byłoby nie wiedzieć?

### CZY LEPIEJ NIE WIEDZIEĆ?

Statystyka publiczna, która gromadzi dane z wielu dziedzin życia społecznego i gospodarczego, zna takie zmienne i takie przekroje badanych zbiorowości, o które nie pyta lub o których nie informuje. Dziesięć lat temu francuska federa-

cja CRAN (Conseil Représentatif des Associations Noires), zrzeszająca organizację osób czarnoskórych, zleciła badanie na temat traktowania mniejszości etnicznych we Francji. Okazało się ono przełomowe, nikt wcześniej bowiem nie dzielił obywateli — jak podawała prasa<sup>1</sup> — według koloru skóry. Zbieranie tego typu danych było wręcz zakazane. Model republikański, który przyjęła Francja, zakładał, że w kraju nie ma społeczności etnicznych, są tylko obywatele, którzy mają prawa i obowiązki wobec państwa. Również obecnie część danych statystycznych dotyczących np. rasy osób podejrzanych o przestępstwa lub skazanych nie jest w niektórych krajach ujawniana<sup>2</sup>. Są więc w życiu społecznym przesłanki — słuszne lub nie — aby niektórych danych nie rejestrować i nie udostępniać nawet w formie zagregowanej.

O tym, że lepiej nie znać niektórych wskaźników społecznych, pisała też Wisława Szymborska w niezwykłym wierszu pt. *Przyczynek do statystyki* (Szymborska, 2011). Charakteryzując w nim kilkanaście różnych cech i postaw zwykłych ludzi (*Na stu ludzi...*), w jednym przypadku noblistka stwierdza, że lepiej nie znać właściwego odsetka:

(...)  
okrutnych,  
kiedy zmuszą ich okoliczności  
— tego lepiej nie wiedzieć  
nawet w przybliżeniu.

Równie groźnie jak ta strofa wiersza brzmi tytuł wydanej w 2017 r. książki Cathy O’Neil: *Broń matematycznej zagłady. Jak algorytmy zwiększają nierówności i zagrażają demokracji* (*Weapons of Maths Destruction. How Big Data Increases Inequalities and Threatens Democracy*). Czytając ją, można pomyśleć, że wobec tytułowej zagłady chyba lepiej byłoby nie zbierać niektórych danych, nie dawać ich do przetwarzania algorytmom. Może lepiej byłoby nie wiedzieć?

### RZECZYWIŚCIE MATEMATYCZNA BROŃ I ZAGŁADA?

Warto zwrócić uwagę, że pracę O’Neil pozytywnie oceniła recenzentka z SGH prof. dr hab. Leokadia Oręziak, pisząc m.in.: *Książka O’Neil jest praktycznie pierwszą książką na świecie, która w tak obszerny i pogłębiony sposób przedstawia, jak zjawisko Big Data stało się jednym z kluczowych czynników wzrostu nierówności społecznych i zagrożeń demokracji*<sup>3</sup>. Mimo że publikacja warta jest uwagi i dyskusji, za ryzykowne uważam stwierdzenie o jej wyjątkowości w skali świata<sup>4</sup>, a za przesadę nazwanie Big Data czynnikiem kluczowym dla wzrostu nierówności społecznych. O’Neil wini co prawda algorytmy za rozprzestrzenianie się

<sup>1</sup> Czarnoskórzy Francuzi liczą siebie i tolerancję. *Gazeta Wyborcza*, 1.02.2007 r.

<sup>2</sup> Urząd statystyczny Kanady nie zbiera i nie udostępnia danych wymiaru sprawiedliwości w sprawach karnych w przekrojach etnicznych. (Owusu-Bempah i Wortley, 2014). Według podobnej zasady postępuje sztokholmska policja od roku 2016, chcąc uniknąć stygmatyzowania całych grup etnicznych bądź narodowych (Swedish police banned from describing criminals anymore in case they sound racist. *Independent*, 14.01.2016 r.).

<sup>3</sup> Z obwoluty książki.

<sup>4</sup> Wiele zagrożeń związanych z wykorzystywaniem Big Data zostało już opisanych (Mayer-Schönberger i Cukier, 2013).

różnego rodzaju nierówności, ale jednocześnie dostrzega (pisząc w rozdziale 2 m.in. o funduszach hedgingowych) nieetyczne działania finansistów oraz oszustwa firm ratingowych w okresie poprzedzającym kryzys finansowy 2008 r. Podobnie rzecz się ma z wpływem Big Data na demokrację. Można w tej monografii znaleźć niemało argumentów przekonujących o zakłócaniu reguł demokratycznych przez wykorzystanie w marketingu politycznym ogromnych możliwości Big Data. Trudno jednak byłoby wskazać na ich kluczowe znaczenie.

Za przesadzone uważam ponadto użycie przez polskiego tłumacza (Marcina Z. Zielińskiego) słowa *zagłada*, mającego dużo bardziej emocjonalne zabarwienie niż angielskie *distraction* — zniszczenie, zrujnowanie. W całej książce, w ślad za Autorką, tłumacz operuje skrótem *Beemzet* dla określenia broni matematycznej zagłady. Gdyby zamiast słowa *zagłada* użył słowa *zniszczenie*, to skrót pozostałby ten sam, a uniknąłby — motywowanej prawdopodobnie względami marketingowymi — emocjonalnej przesady.

Jeśli zaś chodzi o obecny w polsko- i anglojęzycznym tytule książki przymiotnik *matematycznej*, to od razu pragnę uprzedzić Czytelników, że matematyki jest w całym tym opracowaniu jedynie tyle, ile wykorzystuje się jej w statystyce, metodach symulacyjnych (Monte Carlo) i sztucznych sieciach neuronowych. W rzeczywistości chodzi więc nie o matematykę, ale o statystykę, co dla Autorki nie ma większego znaczenia — pisze zamiennie o matematykach i statystykach.

### ZAGROŻENIA ZWIĄZANE Z BIG DATA

O'Neil trafnie diagnozuje najważniejsze zagrożenia związane z Big Data, chociaż nie jest prekursorką tego tematu. Zwraca uwagę przede wszystkim na to, że jakkolwiek złożone byłyby modele, na podstawie których klasyfikujemy bądź oceniamy ludzi (nauczycieli, kredytobiorców, ubezpieczających się), a także obiekty (uczelnie — rozdział 3), to ich konieczne uproszczenia mogą być powodem niesprawiedliwości i krzywdy człowieka<sup>5</sup>. Zastrzeżenia te Autorka odnosi m.in. do testów oceniających jakość pracy nauczycieli, opartych na tzw. wartości dodanej, a także do rankingów wyższych uczelni. Tu jednak problem wydaje się mniej poważny, bo — choć Autorka tego wyraźnie nie napisała — wystarczy potraktować wyniki użycia algorytmów nie jako jedyne, ale jedno z kilku kryteriów oceny jednostek w zbiorowości.

Poważniejsze obawy budzi u Autorki wykorzystanie algorytmów predyktywnych (np. w walce z przestępczością), a także innych modeli używanych w problematyce penitencjarnej. Przekonująco brzmią zwłaszcza argumenty, że konstrukcja niektórych modeli może prowadzić do wciągnięcia skazanej osoby w mechanizmy, które się nawzajem wzmacniają (tworząc negatywne sprzężenia zwrotne) i praktycznie uniemożliwiają jej powrót do normalnego życia. W publikacji często pojawiają się hipotetyczne bądź rzeczywiste przykłady osób z biednych środowisk oraz reprezentantów czarnoskórej społeczności w Stanach Zjednoczonych.

Równie uzasadnione — mimo że i tym razem nieoryginalne — są niepokoje Autorki o możliwości wykorzystania algorytmów Big Data do nierównego traktowania klientów przez sprzedawców dóbr lub usług. Wgląd w przeszłość klienta i jego domniemane dochody stwarza poważne niebezpieczeństwo potencjalnych

<sup>5</sup> Autorka w swojej wrażliwości społecznej bliska jest Sandelowi (2013).

nadużyć dokonywanych przez wszystkich, którzy będą mieli dostęp do tych odpowiednio przefiltrowanych i przeanalizowanych danych.

Wiele postulatów zawartych w książce O'Neil wartych jest uwagi. Za najistotniejszy uważam potrzebę transparentności algorytmów, którymi posługują się instytucje i przedsiębiorstwa, które mają najszerzy dostęp do naszych danych i do możliwości ich przetwarzania. Autorka proponuje, aby środowisko akademickie angażowało się na rzecz stworzenia systemu audytów modeli i algorytmów Big Data. W algorytmach nieaktualizowanych, źle skonstruowanych, celowo krzywdzących słabsze jednostki lub nakierowanych na optymalizację w zbyt krótkiej perspektywie Autorka dostrzega źródło najpoważniejszych zagrożeń. Trudno się z tym nie zgodzić. Nie jesteśmy więc, jak sądzę, skazani na zagładę — ani jako uczestnicy życia gospodarczego, ani społecznego, ani politycznego. Stoimy przed nowymi wyzwaniami, w sprostaniu którym największego wsparcia oczekiwać należy właśnie od matematyki, a ściślej — informatyki i statystyki.

---

**prof. dr hab. Mirosław Szreder** — *Uniwersytet Gdański*

#### LITERATURA

- Mayer-Schönberger, V., Cukier, K. (2013). *Big Data. Rewolucja, która zmieni nasze myślenie, pracę i życie*. Warszawa: Wydawnictwo MT Biznes.
- Owusu-Bempah, A., Wortley, S. (2014). Race, Crime, and Criminal Justice in Canada. W: S. Bucerius, M. Tonry (red.), *The Oxford Handbook of Ethnicity, Crime, and Immigration* (s. 281—320). Oxford University Press.
- Sandel, M. (2013). *Czego nie można kupić za pieniądze*. Warszawa: Kurhaus.
- Szyborska, W. (2011). *Przyczynek do statystyki*. W: *Chwila/Moment*. Kraków: Znak.