

R E C E N Z J E , O M Ó W I E N I A

ANGELINA KUSOWSKA
Uniwersytet Łódzki

JAK DOBRZE ZNASZ SWOJEGO DATA SCIENTISTĘ?

W *Data science: najseksowniejszy zawód XXI wieku w Polsce* Remigiusza Żulickiego¹ czytamy: „Data science to dynamiczne, zmienne, nowe, bardzo słabo rozpoznane i niejasne zjawisko społeczne o poważnych konsekwencjach dla innych sfer życia” (s. 9). Takie krótkie określenie nie rozjaśnia głównego przedmiotu badań i sugeruje, że badacz zajmuje się obszarem mało znanym, niełatwym do analizy i trudnym do zweryfikowania na poziomie ostatecznej interpretacji wyników. Autor sam, z założenia, poznaje tajemnice świata społecznego data scientistów — osób wykonujących operacje na danych cyfrowych z wykorzystaniem skryptowych języków programowania (s. 9) — testuje

i eksperymentuje w zakresie metodologii, próbując przełożyć techniki studiów jakościowych na świat zdominowany przez technologię.

Czy przedstawione badania rzeczywiście pozwalają nam dowiedzieć się czegoś nowego o świecie społecznym data scientistów?

Po lekturze książki odpowiedź na jej tytułowe przesłanie nie jest jednoznaczna. Społeczny obraz data scientistów nie został przez Żulickiego zmieniony, a sami data scientiści nie stali się w naszych oczach „superbohaterami kodu” (nie to było celem badacza), ale z pewnością dostaliśmy obraz precyzyjnie namalowany i omówiony w najdrobniejszych szczegółach — a to wciąga i inspiruje.

Około trzystu stron dogłębnej analizy jakościowej, którą otrzymuje czytelnik, to zaledwie wierzchołek góry lodowej, jaką jest świat społeczny *data science* (DS). Możemy odnaleźć potwierdzenie tych słów po rozłożeniu skomplikowanej (nie tylko na pierwszy rzut oka) karty-mapy świata DS w Polsce (s. 232), która ukazuje najważniejsze elementy szczegółowo ujętej areny społecznej. Mapa, choć skutecznie przy-

Adres do korespondencji: angelakusowska@gmail.com; ORCID: 0009-0008-4137-9498

¹Remigiusz Żulicki, *Data science: najseksowniejszy zawód XXI wieku w Polsce. Big Data, sztuczna inteligencja i PowerPoint*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2022, stron 348 (numery cytowanych stron w tekście, w nawiasach).

ciąga uwagę czytelnika, stanowi zaledwie zapowiedź wniosków autora, skłaniających do podjęcia refleksji w zakresie szeroko pojętego DS jako paranauki, ale również samego zjawiska związanego z postępującą technologizacją społeczeństwa.

Według autora główny cel jego pracy stanowi: „opis etnograficzny polskiego społecznego świata data science”. Jest to pierwsza socjologiczna monografia, która opisuje DS w kategoriach świata społecznego w rozumieniu Adele E. Clarke. Obszerna książka powstała na podstawie rozprawy doktorskiej „Data Science w Polsce. Etnografia społecznego świata”, przedstawionej i obronionej z wyróżnieniem w 2021 roku, a rok później w konkursie im. Aleksandra Matejki uznanej za najlepszą rozprawę doktorską z zakresu socjologii pracy.

Rozważania teoretyczne Żulicki rozpoczyna od przybliżenia czytelnikowi poszczególnych terminów związanych z dziedziną DS. Dzięki temu wiemy, że *big data* są „wielkie”, „różnorodne” i „szybkie” (s. 11), a *AI* „stanowi dla laika pokaz magiczny” (s. 15), chociaż w rzeczywistości wcale nie istnieje: „pojęcie *AI* jest z punktu widzenia uczestników świata DS tak wieloznaczne i budzące emocje, że w sensie technicznym nie da się zaakceptować jego użycia” (s. 287). Na pierwszych kilkudziesięciu stronach nie brakuje odniesień do klasyków socjologii (Anselm Strauss, Howard Becker), a także współczesnych przedstawicieli tej dziedziny w Polsce (Anna Kacperczyk, Kazimierz Krzysztofek, Krzysztof Konecki, Dariusz Jelmieniak). Żulicki korzysta z ustaleń takich badaczy jak: Kenneth Cukier, Cathy O’Neil, Aleksandra Przeglalińska, Przemysław Biecek. Najważniejsze miejsce zajmuje jednak Adele E. Clarke, autorka analizy sytuacyjnej oraz strategii ugruntowanego teoretyzowania.

Przywoływane przez autora osobistości ze świata nauki (zarówno tego technicznego, jak i społeczno-humanistycznego) — wskazują, że jako badacz do omawianego problemu podszedł w sposób otwarty i interdyscyplinary. Dla kształtu monografii duże znaczenie ma

także podłoże autoetnograficzne (*background*; s. 61), przy czym Żulicki opowiada też o swoich dziecięcych fascynacjach oraz wczesnych zainteresowaniach obszarem nauk matematycznych. Podejście to, zwłaszcza w obliczu przeprowadzonej przez badacza analizy, wydaje się rzeczywiście rozwojowe dla świata nauki, a także odkrywcze — w sposób dosłowny, gdyż autor nie tylko wyjaśnił tajemniczą definicję „data science”, ale także dokonał kompleksowego odkrycia... czy też „odstąpienia” nieusystematyzowanych do tej pory faktów o uczestnikach świata społecznego DS w Polsce.

Rozmiar się (nie) liczy

Mimo że „prawdziwa” sztuczna inteligencja, zdaniem badacza oraz innych uczestników świata DS, jeszcze nie istnieje, to temat ten w zestawieniu z *big data*, *data science* oraz uczeniem maszynowym cieszy się największą popularnością (s. 17). Żulicki wielokrotnie zwraca uwagę na fakt, który winien mieć istotne znaczenie dla uczestników świata społecznego DS: „Sztuczna inteligencja to tylko kolorowe opakowanie dla różnych systemów, które mogą, ale nie muszą działać na bazie *ML*. [...] Nie ma w badanym świecie zgody co do tego, czy używanie rozmytego, obrośniętego popkulturowymi skojarzeniami terminu *AI* w komunikacji ze światami nietechnicznymi leży w interesie świata DS” (s. 253). Autor tłumaczy „magię” *AI*, powołując się na praktyki celowo stosowane przez biznesowych rzeczników oraz na powszechnie znane w świecie DS powiedzenie, że sztuczna inteligencja „napisana jest w *PowerPoint*” (s. 251). W rzeczywistości *data science* nie „czarują” i nie tworzą obdarzonych cyfrową świadomością maszyn, potrafią jednak „masterkować”, stawiać i sprawdzać własne hipotezy czy porządkować dane w taki sposób, by znalazły one zastosowanie w realnych problemach rzeczywistego świata. I chociaż stworzone przez DS narzędzia obarczone są ciężarem odpowiedzialności za bezpieczeństwo ich użytkowników, to jednak rozwiązania te nie są z natury ani dobre, ani złe, a prawdopodobieństwo

utworzenia maszyny dorównującej inteligencją człowiekowi jest obecnie zbyt małe, by rozważać jej rzeczywiste zastosowanie: „Technologie, w tym tzw. AI czy — technicznie rzecz ujmując — ML, postrzegane są raczej jako neutralne narzędzia w ludzkich rękach” (s. 41).

Badacz celowo pomija problematykę „złego” i „dobrego” użycia AI, przyznając się do tego kilkakrotnie i wspominając zaledwie w kilku słowach o najgłośniejszych aferach ze sztuczną inteligencją w roli głównej. Choć dociera do nas argument, zgodnie z którym obszerne informacje na ten temat można znaleźć na własną rękę (np. o Cambridge Analytica), a sama praca nie miała dotyczyć sztucznej inteligencji oraz konsekwencji jej rozwoju, to jednak brakuje czytelnikowi podrozdziału w pełni poświęconego pozytywnym i negatywnym wpływom DS. Zwłaszcza gdy książka sugeruje, że część stron poświęconych została big data czy słynnemu AI (tytuł książki), a pytania retoryczne stanowiące opis data science dotyczą kojarzonych z *clickbaitami* wątpliwości (Czy sztuczna inteligencja pozabawia nas pracy? Algorytmy przejmują władzę nad światem?). Analiza przeprowadzona przez Żulickiego miała się skupić głównie na ludzkich i nie-ludzkich uczestnikach świata społecznego DS, co udało się osiągnąć — żal jednak że, nie udało się skorzystać z okazji do rozmowy z samymi data scientistami, by poznać lepiej zakres ich społecznej odpowiedzialności.

Znacznie więcej uwagi autor poświęca zagadnieniom rozwoju technologicznego w ogóle, z wyszczególnieniem mniej znanych w świecie nietechnicznym narzędzi. Data science to przede wszystkim analiza danych, tworzenie modeli statystycznych, które są zdolne do uczenia się zachodzących prawidłowości. Na kodach stworzonych przez data scientistów rozwijają się między innymi wszelkie systemy rekomendacji, inteligentne wyszukiwanie, mechanizmy antyspamowe czy chatboty. Żulicki bardzo dokładnie definiuje DS, oddzielając je od świata programowania: „Działanie podstawowe społecznego świata DS w Polsce określam jako pisanie kodu do przetwarzania, analizy i modelowania da-

nych” (s. 76). Podobnie data science rozumie Thomas H. Davenport w artykule *Data Scientist: The Sexiest Job of the 21st Century*², na którego ustalenia autor powołuje się w pracy.

Zdaje się, że według opinii publicznej pojęcia związane z technologiami data science są pociągające, odświeżające, egzotyczne, a co najważniejsze — obiecujące. Brak fundamentalnej wiedzy technicznej wśród odbiorców komunikatów biznesowych i prasowych, które dotyczą data science, wywołuje w konsumentach pragnienie dokonania niemożliwego z pomocą doświadczonych „grotołazów danych” (*data spelunker*), „wygłodniałych badaczy”, „sceptycznych detektywów” czy „ludzi renesansu” (s. 22). Osoby spoza świata technicznego mogą być zaskoczone tym, że inteligentne modele wcale nie odbiorą nam stanowisk pracy.

Na styku techniki i socjologii

W toku lektury *Data science: najseksowniejszy zawód świata* czytelnik jest zapoznawany z dużą, niekiedy przytłaczającą liczbą definicji, które należy umiejętnie rozróżniać, by w pełni zrozumieć strukturę świata społecznego DS. Dla laika, który dopiero uczy się rozpoznawania tego świata, stos obcojęzycznych nazw systemów wykorzystywanych powszechnie przez data scientistów może wzbudzać niepokój. Być może kilkakrotnie będzie musiał zatrzymać się na wybranej stronie, by we własnym zakresie ułożyć sobie w głowie poszczególne dotychczas nieznanne definicje. Mnogość trudnych do wypowiedzenia i jeszcze trudniejszych do zapamiętania nazw narzędzi, którym przypisane zostały pewne funkcje w świecie DS, niestety może także uspić czujność czytelnika, w szczególności tego nietechnicznego. Gdy podczas lektury napotkałam kilka stron tekstu nasyconych technicznym żargonem lub rozwiniętymi analizami zebranych danych ilościowych,

²Thomas H. Davenport, D. J. Patil, *Data Scientist: The Sexiest Job of the 21st Century*, „Harvard Business Review” 2012, t. 90(10), s. 70–76.

mimo największych starań nie udało się w pełni skupić na sensie niektórych przedstawionych wniosków. Skomplikowane tłumaczenia technicznych struktur, w których uczestnicy świata DS zazwyczaj się poruszają, momentami sprawiały, że lektura stawała się trudna i powolna. Jednak po każdej definicyjnej fali następował swoisty odpływ — odkrywający szczególnie interesujące uwagi, znacznie łatwiejsze do zrozumienia dla nietechnicznego laika, tłumaczące w skrócie poprzedzające je stronicę.

Żulicki zacytował definicję DS, którą zaproponowała Grażyna Trzpiot³: „Data science jest procesem zamieniania/przekształcania danych w konkretne działania”. Statystyka ma zastosowanie w jednym z czterech obszarów działań DS. Są to: „uzyskaj» — pozyskanie danych; «przygotuj» — przygotowanie danych; «analizuj» — tutaj wymagana jest znajomość statystyki i innych metod analizy danych; «działaj» — wdrożenie wyników (Trzpiot, 2017: 11–15)” (s. 31). Nie da się oprzeć wrażeniu, iż autor książki, podobnie jak data scientist, podjął próbę napisania swoistego „kodu” świata społecznego DS. Pozyskując dane podczas kilkuletnich badań etnograficznych; przygotowując dane poprzez spójne połączenie informacji zebranych „osobno i bez założenia, że będą porównywalne” (s. 58); analizując dane, z wykorzystaniem powszechnych dla uczestników świata DS narzędzi do pozyskiwania, przetwarzania i modelowania danych. Tak „napisany kod” stanowi podstawę pierwszej obszernej polskiej monografii o data scientistach przygotowanej przez socjologa. *Data science: najseksowniejszy zawód...* to przykład socjologicznego, jakościowego pochylenia się nad obszarem nauk technicznych, pokazujący, iż za skomplikowane procesy data science odpowiadają nie inteligentne algorytmy zaprojektowane przez anonimowych ludzi, lecz osoby o pewnych cechach oraz przyzwyczajeniach,

przejawiające określone tendencje, które łączą przedstawicieli tej grupy zawodowej, umożliwiając im wytworzenie swoistej wspólnoty, którą można zbadać oraz opisać.

Lektura *Data science: najseksowniejszy zawód XXI wieku w Polsce. Big data, sztuczna inteligencja i PowerPoint* skłania do refleksji nie tylko nad światem społecznym DS, ale również nad zagrożeniami cyfrowymi w społeczeństwie technologicznym w ogóle. Zdaje się jednak, że problemy te warto byłoby poruszyć nieco szerzej, być może w osobnym rozdziale — dzięki czemu książka nie tylko opisywałaby świat społeczny DS, ale również uświadamiałaby społeczną rolę data scientistów. Zrozumienie wpływu DS na rozwój współczesnych zjawisk społecznych uważałabym za adekwatne do opisu areny społecznej zaprezentowanej przez Żulickiego. Obszerniej wyjaśniłoby to relacje zachodzące pomiędzy poszczególnymi grupami znajdującymi się na przygotowanej przez autora mapie (s. 232), tym samym być może precyzując mechanizmy integracji między DS a resztą świata.

Recenzowana książka (tak jak inne możliwe monografie socjologiczne o DS) może z jednej strony pomóc w rozwoju świata DS (zrozumienie potrzeb, predyspozycji i wewnętrznych zasad oraz dostosowanie nowych rozwiązań ułatwiających funkcjonowanie DS), z drugiej może stwarzać zagrożenie dla naturalnej skrytości działań data scientistów. Chociaż z pewnością rozumieją oni powszechne zagrożenia związane z dezinformacją w środowisku cyfrowym i nierzadko podejmują działania zapobiegające umieszczeniu ich w algorytmicznych bańkach filtrujących, to nie można nie mieć obaw, że badania jakościowe nad obszarem data science mogą to utrudnić. Odsłaniając życie data scientistów, nauka wystawia przedstawicieli tego środowiska na ataki ze strony specjalistów od „targetowania” pracujących nad tym, by dotrzeć do jak największej liczby osób obdarzonych zdolnościami technicznymi. Gdzie znajdują się data scientiści? Czego szukają w sieci? Jak spędzają swój czas wolny? Jakie mają wartości i co może wzbudzić ich ciekawość? Nale-

³Grażyna Trzpiot, *Rozumienie Data Science*, w: Grażyna Trzpiot (red.), *Statystyka a Data Science*, Wydawnictwo Uniwersytetu Ekonomicznego, Katowice 2017, s. 9.

ży pamiętać, że odpowiedzi na te pytania, choć są cenne dla współczesnej socjologii opisującej nowoczesne społeczeństwo technologiczne, mogą również sprzyjać strategiom charakterystycznym dla inżynierii społecznej wycelowanej w tych, którzy „czynią magię”. Być może te uwagi wykraczają daleko poza tematykę podjętą przez Żulickiego. Sądzę jednak, że tak ważna publikacja, jaką jest pierwsza polska monografia socjologiczna o DS, powinna dokładniej wymienić i omówić zagrożenia towarzyszące światu data science.

Społeczny świat DS — wnioski z lektury

Chociaż w obszernym tekście zabrakło, moim zdaniem, miejsca na bardzo istotne zagadnienia związane z bezpieczeństwem oraz scenariuszami rozwoju data science jako dość skrytej i zmiennej dziedziny, a informacyjny rollercoaster chwilami uniemożliwia płynne śledzenie poruszanych przez autora wątków, to przyznać należy, że o sukcesie dydaktycznym tej książki świadczy fakt, iż pozostawia czytelnika z inspiracjami oraz refleksjami, które chciałoby się skonfrontować z kolejnymi badaniami DS. Czytając kolejne rozdziały książki, można zauważyć, że przedstawiane wyniki badania nie tylko opisują (konstruuja) świat DS, ale także odsłaniają to, co intymne. Obserwacje uczestniczące, notki autoetnograficzne oraz fragmenty wywiadów swobodnych przeprowadzonych z uczestnikami badań pozwalają przyjrzeć się prywatnemu życiu data scientistów. Ich emocjom, dotychczasowym doświadczeniom, opiniom, relacjom, rutynie, komunikacji. Można by nawet uznać, że opisywana monografia „rozbiera do naga” arenę społeczną DS. Część analizy autor poświęca bowiem między innymi stylom ubierania się, sposobom wypowiedzi, gustom, cechom osobowości i temperamentom data scientistów: „Latem w pełni akceptowane są osoby noszące krótkie spodnie, sandały lub klapki” (s. 192).

Wszystkie elementy, ludzkie oraz nieludzkie, składające się na świat społeczny DS przed-

stawione zostały za pomocą grafu uwzględniającego relacje między areną DS a innymi światami społecznymi (przedstawiciele różnych stanowisk, sojusznicy, wrogowie — s. 230). Podsumowując swoje wnioski na temat świata społecznego DS, w opisie głównej areny sporów badacz postanowił uchwycić pełną sytuację badania. Graf jest na tyle skomplikowany, że niełatwo jest zrozumieć tę kompleksowo przedstawianą arenę. Żulicki nie pozostawia jednak czytelnika bez merytorycznego wsparcia, dokładnie tłumaczy swoje obserwacje. Utworzoną mapę oraz dogłębną analizę uzupełniają wypowiedzi uczestników badania, osób należących do świata data science — do tych uwag autor odnosi się w trakcie swojego wyводу.

Data science: najseksowniejszy zawód XXI wieku w Polsce. Big data, sztuczna inteligencja i PowerPoint to książka, która może zainteresować nie tylko socjologów zafascynowanych nowoczesnymi technologiami, ale również osoby „niesocjologiczne” (chcące zrozumieć funkcjonowanie społeczności data science) lub „wannabesów” (pragnących do tej społeczności dołączyć). Przeprowadzone badania ukazują istotę realnego uczestnictwa socjologów w procesach społecznych badanego środowiska. Jeśli socjolog chce poznać i opisać wybraną arenę społeczną, powinien w nią wejść — powinien ją rozłożyć na części, zrozumieć od środka oraz poczuć ją tak jak r z e c z y w i ś c i uczestnicy tego świata społecznego. Przedstawiona analiza jakościowa świata DS może ułatwić socjologom późniejsze zrozumienie nowych technologii na poziomie ich wytwarzania, a nie jedynie określenia ich konsekwencji czy funkcji.

Data science: najseksowniejszy zawód XXI wieku w Polsce. Big data, sztuczna inteligencja i PowerPoint swoją złożonością i głębią analiz przywoździ na myśl przyrodnicze programy dokumentalne. W przypadku książki Remigiusza Żulickiego „lektorem” jest sam czytelnik, a informacje nie dotyczą tajemniczego gatunku zwierząt czy roślin, lecz wybranej zbiorowości zawodowej. W obu sytuacjach pozostaje jednak pytanie — jak ową wiedzę wykorzystać?