Marta Daćków

Automatyzacja procesów decyzyjnych a problem dyskryminacji w kontekście regulacji europejskich

Spis treści

Wstęp ... 3
1. Przedmiot pracy i pytania badawcze ... 3
2. Teza ... 7
3. Metodologia pracy ... 7
4. Struktura pracy ... 10
I. Różnicowanie a dyskryminacja. Historyczne ujęcie problemu ... 14
1. Historyczna walka o równość ... 16
2. Dyskryminacja w perspektywie społecznej ... 25
3. Pojęcie dyskryminacji ... 32
4. Formy dyskryminacji ... 37
II. Sztuczna inteligencja i jej wpływ na rozwój automatycznych systemów decyzyjnych ... 44
1. Automatyczne systemy decyzyjne ... 44
2. Pojęcie sztucznej inteligencji, uczenia maszynowego i *big data* ... 53
3. Różnicowanie przez automatyczne systemy decyzyjne a problem automatycznej dyskryminacji ... 63
III. Europejskie regulacje prawne jako źródło zakazu różnicowania w algorytmicznych procesach decyzyjnych ... 74
1. Historia i treść prawnego zakazu dyskryminacji ... 75
2. Europejskie prawo antydyskryminacyjne jako podstawa zakazu różnicowania przez ASD w świetle orzecznictwa sądów ... 81
3. Ochrona danych osobowych ... 94
4. Prawo do prywatności ... 109
IV. Sztuczna inteligencja jako samodzielny przedmiot regulacji prawa Unii Europejskiej ... 119
1. Droga do uregulowania sztucznej inteligencji ... 120
2. Akt w sprawie sztucznej inteligencji ... 127
3. Logika regulacji SI ... 150
V. Automatyczna dyskryminacja ... 159
1. Redefiniowanie zakresu i form dyskryminacji ... 160
2. Ryzyko jako podstawa regulacji ... 174
3. W kierunku materialnej równości ... 179
Zakończenie ... 185
Bibliografia ... 188

2:11272

Wstęp

1. Przedmiot pracy i pytania badawcze

Zagrożenia związane ze stosowaniem automatycznych systemów decyzyjnych, w tym algorytmów sztucznej inteligencji, a także wpływ cyfryzacji na jednostki i społeczeństwo to jedne z najistotniejszych zagadnień współczesnego świata, a więc również prawa. Chociaż modele przewidywania ryzyka stosowano wiele lat przed rozpowszechnieniem sztucznej inteligencji, aktualnie obserwujemy dynamiczny rozwój rynku automatyzacji decyzji, który funkcjonuje na styku technologii, społeczeństwa i prawa. Systemy te, łudząc użytkowników technologiczną neutralnością, z jednej strony są statystycznym odzwierciedleniem rzeczywistości – przez co powielają i ugruntowują procesy różnicujące istniejące w społeczeństwach – z drugiej zaś kreują nowy problem nieprzejrzystego klasyfikowania ludzi. Pojawiające się przy tym ryzyko naruszenia praw i wolności jednostki ma charakter masowy oraz inherentny i obejmuje zarówno sferę prywatną, jak i publiczną funkcjonowania człowieka w społeczeństwie.

Cywilizacyjne skutki upowszechnienia automatyzacji procesu decyzyjnego mają wymiar polityczny i komercyjny. Przykładowo w Chinach publiczny system zaufania społecznego (ang. *Social Credit System*) stale monitoruje i ocenia zachowania wszystkich obywateli, dokonując ich zaszeregowania do odpowiednich grup. System opiera się na założeniach znanych w Europie z mechanizmów scoringów kredytowych, jednak sięga dużo głębiej, uderzając w prywatność obywateli. Dane pozyskiwane są bowiem nie tylko od podmiotów publicznych (jak np. rejestry karne), ale też podmiotów prywatnych (np. od Sesame Credit, będącej instytucją finansową powiązaną z portalem Alibaba). Skutkami zakwalifikowania jednostki do konkretnej grupy są: przydział świadczeń publicznych, dostęp do rynku, w tym mieszkaniowego i kredytowego, a także przyznanie prawa wyjazdu z kraju. W uproszczeniu jest to mechanizm kar i nagród o wymiarze publicznoprawnym, w którym o prawach jednostki decyduje algorytm opierający się na analizie ogromnej ilości danych[1].

Głośnym echem w mediach odbiły się również rozwiązania przyjmowane przez firmę Amazon. Wprowadzony tam automatyczny system rekrutacji dyskryminował bowiem kobiety, a dokładniej – odrzucał ich CV na podstawie danych pośrednio świadczących o płci kandydata

[1] C. Hatton, *China 'Social Credit': Beijing Sets Up Huge System*, BBC 2015, https://www.bbc.com/news/world-asia-china-34592186 (dostęp: 4.12.2021).

3:34462

(np. ukończona szkoła żeńska), mimo że był opracowany tak, aby być neutralny płciowo (system był tworzony na podstawie danych Amazona sprzed 10 lat). Problem był jednak na tyle złożony, że ostatecznie firma, nawet dysponując wiedzą o tych błędach, w obawie przed innymi przejawami dyskryminacji zrezygnowała z wykorzystywania narzędzia[2]. To niejedyny przykład dyskryminującego algorytmu stosowanego przez Amazona. Kolejnym był system uniemożliwiający bezpłatną dostawę do dzielnic zamieszkiwanych przez mniejszości etniczne[3].

Jako przykład naruszenia na szeroką skalę podawany jest też amerykański system PredPol, służący do predykcji przestępczości. Doprowadził on bowiem do niezamierzonego skoncentrowania działań funkcjonariuszy policji na drobnej przestępczości w ubogich dzielnicach. Na skutek sprzężenia zwrotnego, polegającego na potwierdzeniu skuteczności wykrywania przestępczości mniejszej wagi (np. włóczęgostwo, używanie niewielkich ilości narkotyków), nieproporcjonalnie karano osoby z mniejszości narodowych zamieszkujących takie dzielnice, podczas gdy przestępstwa z użyciem przemocy nie były lokalizowane w systemie.

Podobny efekt sprzężenia zwrotnego osiągnął amerykański system COMPAS, zaprojektowany w celu algorytmicznego wspomagania neutralności wyroków. Stosowanie systemu doprowadziło tymczasem do umocnienia, a nie wyeliminowania uprzedzeń na tle rasowym[4]. Równie krytycznie oceniane były błędy algorytmiczne powiązane z płcią i kolorem skóry w technologii rozpoznawania twarzy – system Microsoft generował wielokrotnie częściej fałszywie pozytywne wyniki w przypadku czarnoskórych kobiet niż czarnoskórych mężczyzn. Narzędzie to w USA doprowadziło przez to w 2020 r. do niesłusznego zatrzymania i aresztowania Afroamerykanina[5]. Wiadomo było przy tym, że system używany przez amerykański wymiar sprawiedliwości miał wysoki wskaźnik błędnej identyfikacji osób czarnoskórych, jednak mimo żądania ograniczenia stosowania budzącego kontrowersje narzędzia predykcyjnego projekt zaprzestania jego wykorzystania został odrzucony w 2021 r.

[2] J. Dastin, *Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women*, Reuters 2018, https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-airecruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G (dostęp: 4.12.2021).

[3] D. Ingold, S. Soper, *Amazon Doesn't Consider the Race of Its Customers. Should It?*, Bloomberg 2016, https://www.bloomberg.com/graphics/2016-amazon-same-day/?cmpid=google (dostęp: 4.12.2021).

[4] C. O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, New York 2017.

[5] K. Hill, *Wrongfully Accused by an Algorithm*, „New York Times" 2020, https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html (dostęp: 4.12.2021).

4:95910

Opisane mechanizmy związane są głównie z błędami w analizie danych, m.in. poprzez nadawanie wysokiej wagi bezpośrednio lub pośrednio cechom chronionym prawem antydyskryminacyjnym, jak rasa, płeć czy pochodzenie etniczne, jak również poprzez oparcie analizy na niereprezentatywnych zbiorach danych. Działanie zaawansowanych algorytmów może być jednak o wiele mniej oczywiste, ponieważ klasyfikowanie ludzi przez system może się odbywać na podstawie informacji czy cech niepodlegających obiektywnej ocenie[6] lub danych i/lub informacji niestanowiących cech chronionych w rozumieniu prawa antydyskryminacyjnego, a nawet cech ludzkich (np. dane behawioralne).

Aktualnie największe zainteresowanie budzi ChatGPT[7] – oprogramowanie, które wykorzystuje język naturalny w sposób, który praktycznie uniemożliwia odróżnienie wiadomości generowanych przez algorytm od wypowiedzi pochodzących od człowieka. ChatGPT potrafi przy tym być kreatywny, tworzyć poezję, teksty literackie, a nawet zadania logiczne czy całe programy komputerowe. Algorytm ChatGPT od 2023 r. miał nawet umożliwić generowanie pism prawniczych i wniosków w sprawach reklamacyjnych.

Widzimy zatem, że masowe wykorzystanie algorytmicznych systemów decyzyjnych może nieść za sobą skutki dyskryminacyjne. Naturalne jest więc postawienie pytania o to, na ile istniejące ramy prawne są w stanie odpowiadać na, eliminować lub minimalizować takie naruszenia. Należy przy tym pamiętać, że instytucje prawne uformowane w okresie społeczeństwa nowożytnego, w tym kształt zakazu dyskryminacji, opierający się na cechach chronionych prawem i ludzkich procesach decyzyjnych, jest nieidealną odpowiedzią na postulaty równościowe, w tym fragmentaryczną odpowiedzią na niebezpieczeństwa związane ze współczesną, ponowoczesną ewolucją społeczną, łączącą się m.in. z decentralizacją władzy i jej rozproszeniem na podmioty rynkowe, wykorzystujące zautomatyzowane mechanizmy decyzyjne napędzane zaawansowaną analizą danych. Przeciwdziałanie dyskryminacji poprzez zahamowanie procesów wykluczenia grup historycznie marginalizowanych pozostaje zaś jednym z podstawowych zagadnień społecznych w demokratycznym państwie prawa.

Przedmiotem niniejszej pracy jest analiza instrumentów prawnych prawa europejskiego i zakresu ich zastosowania wobec zaobserwowanego problemu dyskryminacji automatycznej. Wybór tego zagadnienia jest uzasadniony przynajmniej w dwójnasób. Po pierwsze, regulacje

[6] Baskijski system EPV-R (ang. *Intimate Partner Femicide and Severe Violence Assessment*), służący do przewidywania przemocy na tle płciowym, w kwestionariuszu zawiera pytania dotyczące poziomu zazdrości partnera w skali od 0 do 3.
[7] https://openai.com/blog/chatgpt (dostęp: 4.12.2021).

5:33487

te, szczególnie prawodawstwo Unii Europejskiej, są obecnie najszybciej rozwijającymi się ramami prawnymi zarówno dla najbardziej zaawansowanych algorytmów sztucznej inteligencji[8], jak i wykorzystywanych powszechnie procesów decyzyjnych opartych na automatycznym przetwarzaniu danych[9]. Po drugie, prawo europejskie jest ugruntowanym nośnikiem wartości demokratycznych i obywatelskich, wyrażanych m.in. poprzez zakaz dyskryminacji wprowadzony Konwencją o ochronie praw człowieka i podstawowych wolności[10] czy Kartą Praw Podstawowych[11]. Prawo europejskie nie tylko więc coraz mocniej reguluje automatyczne systemy decyzyjne, ale robi to z uwzględnieniem zasad równości i zakazu dyskryminacji.

Poszczególne główne (oznaczone cyframi) i pomocnicze (wyrażone literami) pytania badawcze pracy przedstawiają się następująco:

1) Czy oraz jak automatyczne systemy decyzyjne mogą prowadzić do dyskryminacji?
   a) Czym jest dyskryminacja?
   b) Jak działają i różnicują sytuacje jednostek automatyczne systemy decyzyjne?
   c) Pod jakimi warunkami takie różnicowanie prowadzi do dyskryminacji?
2) Jakie dotychczasowe regulacje prawne mogą w takich sytuacjach znaleźć zastosowanie?
   a) Jakie regulacje prawne odnoszą się do automatycznych systemów decyzyjnych?
   b) Jakim pojęciem dyskryminacji się posługują?
   c) Czy regulacje te obejmują różnicowanie automatycznych systemów decyzyjnych?
3) Jakimi zasadami kieruje się obecna działalność regulacyjna automatycznych systemów decyzyjnych?
   a) Jakie regulacje są projektowane?
   b) Jakie zasady im przewodzą?
   c) Czy zasady te korespondują z dyskryminującym różnicowaniem automatycznych systemów decyzyjnych?

[8] Projekt Rozporządzenia Parlamentu Europejskiego i Rady ustanawiającego zharmonizowane przepisy dotyczące Sztucznej Inteligencji (Akt w Sprawie Sztucznej Inteligencji), COM(2021) 206 final, procedura 2021/0106/COD (dalej: AIA).
[9] Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2016/679 z 27.04.2016 r. w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46/WE, Dz.Urz. UE L 119 (dalej: RODO).
[10] Konwencja o ochronie praw człowieka i podstawowych wolności sporządzona w Rzymie 4.11.1950 r., zmieniona następnie Protokołami nr 3, 5 i 8 oraz uzupełniona Protokołem nr 2, Dz.U. z 1993 r. Nr 61, poz. 284 z późn. zm. (dalej: EKPCz).
[11] Karta praw podstawowych Unii Europejskiej, Dz.Urz. UE C 202 z 7.06.2016 r., 12012P/TXT (dalej: KPP).

6:94892

d) Czy i w jakich przypadkach różnicowanie to może zostać zdefiniowane jako nowa forma dyskryminacji, czyli dyskryminacja automatyczna?

2. Teza

Przeprowadzona analiza prowadzi do wniosku, że automatyczne systemy decyzyjne (dalej: ASD), do których zaliczam algorytmy sztucznej inteligencji, decydując o prawach i obowiązkach człowieka, różnicują jednostki w sposób, który może prowadzić do dyskryminacji. Zakładam, że ASD mogą w związku z tym pogłębiać nierówności społeczne, zwłaszcza powodując na masową skalę wykluczenie grup historycznie marginalizowanych, przy czym procesy różnicowania algorytmicznego nigdy nie będą *explicite* objęte zakazem dyskryminacji w rozumieniu europejskiego prawa antydyskryminacyjnego. Wynika to z charakteru procesu różnicowania w tych systemach, zakładającego niejawność, nieprzejrzystość, masowość, złożoność analizy dużych zbiorów danych i procesów uczenia maszynowego oraz oparcie kategoryzacji na cechach niechronionych prawem antydyskryminacyjnym.

3. Metodologia pracy

Wobec rosnącej specjalizacji i fragmentacji dziedzin nauki interdyscyplinarność badań staje się oczekiwanym postulatem pozwalającym na głębszą analizę problemu, trafniejsze wnioski, jak również recepcję osiągnięć z różnych dyscyplin[12]. Przedmiot niniejszej pracy oraz postawione pytania badawcze wymagają porównania realnych technologicznych procesów zachodzących w rzeczywistości społecznej (automatyczne systemy decyzyjne) ze sferą normatywną (zakaz dyskryminacji w prawie europejskim), dlatego zaplecze metodologiczne opracowania jest interdyscyplinarne, a analizę i badania doktrynalne zostały uzupełnione w nurcie refleksji teoretyczno-prawnej określanej jako „prawo w kontekście" (ang. *law in context*).

Wskazana metoda badawcza została rozwinięta w drugiej połowie XX w. i jest zaliczana do grupy metod zintegrowanych, które odmiennie od podejścia doktrynalnego wychodzą poza

[12] P. Ishwara Bhat, *Idea and Methods of Legal Research*, New Delhi 2019, s. 28.

7:11183

językową analizę i interpretację tekstów prawnych[13]. Podejście to obejmuje teoretyczną i empiryczną analizę prawa jako zjawiska społecznego[14] oraz zakłada analizę problemów prawnych z wykorzystaniem narzędzi i koncepcji charakterystycznych dla socjologii prawa, czyli analizy procesów inżynierii społecznej poprzez prawo i instytucje społeczne. W niniejszej pracy badam zatem prawny problem dyskryminacji oraz walki o równość jako powstały i determinowany przez kontekst społeczny, historyczny i technologiczny. Rozważam zwłaszcza, jak zmiana jednego z tych kontekstów (technologicznego) wpływa na prawne pojęcie dyskryminacji oraz jakie zmiany tego terminu wymusza.

Waga takiej analizy kontekstowej wynika z faktu, że zakaz dyskryminacji jako metaprawo zawsze współistnieje z innymi prawami, w zakresie których postulowana jest równość, w związku z czym norma antydyskryminacyjna jest adaptowana i aktualizowana wraz z rozwojem idei demokratycznego państwa prawnego i poszerzaniem gwarancji prawnych. Integrowanie badań prawnych z innymi obszarami może służyć lepszej analizie wpływu porządkowania społeczeństwa przy wykorzystaniu instrumentów prawa w określonych uwarunkowaniach historycznych. W toku analizy przedmiotu badania, za Davidem Ingramem, przyjmuję, że potrzeba kontekstowego podejścia do praw człowieka związana jest również z uzasadnieniem konieczności ochrony członków społeczeństwa przed marginalizacją i wykluczeniem[15]. Problem dyskryminacji to przy tym zagadnienie zarówno społeczne, jak i prawne, którego znaczenie dla postępującej marginalizacji grup wykluczonych jest zasadnicze.

Realizacja postulatów równościowych zależy od adekwatności gwarancji do aktualnych stosunków władzy i dominacji, te zaś ulegały ciągłym przemianom w kontekstach społeczno-kulturowych oraz technologicznych. Na szczęście interdyscyplinarność badań może pomóc w wyjaśnieniu funkcjonowania prawa w danym porządku społecznym. Z tej przyczyny analizę problemu dyskryminacji w automatycznych systemach decyzyjnych oparłam na kontekście społeczno-historycznym, który pozwala zrozumieć historyczną trajektorię wykształcenia się współczesnego pojęcia dyskryminacji. Drugim kontekstem jest rozwój technologiczny pozwalający na nieznaną wcześniej skalę doprowadzać do różnicowania i kategoryzowania ludzi przez ASD, w tym algorytmy sztucznej inteligencji. W pracy ukazuję, że obecnie to kontekst technologiczny stanowi zaplecze zmian w konstrukcji prawniczego pojęcia dyskryminacji i jej form. Należy przy tym wziąć pod uwagę, że dyskusje na temat sztucznej

[13] *Ibidem*.
[14] *Ibidem*, s. 30.
[15] D. Ingram, *World Crisis and Underdevelopment A Critical Theory of Poverty, Agency, and Coercion*, Cambridge, UK 2018.

8:46330

inteligencji i jej wpływu na świat rozszerzają się z informatyki na inne dyscypliny, w tym prawo. Ciężar i skala zmian, które obejmują nie tylko nowe wymiary podmiotowości, wpływają zatem na podstawowe pojęcia prawa. Te zaś muszą zostać zaadaptowane tak, by jak najlepiej odzwierciedlały interdyscyplinarne przenikanie się zjawisk. Starałam się zatem, aby tło technologiczne nie zostało w pracy ukazane jedynie w świetle teorii ekonomicznej, która widzi sprawiedliwość w kategoriach efektywności, a przez to zaprzecza jakościowym i ewaluacyjnym funkcjom prawa[16].

Odrębność tła wykształcenia się prawa antydyskryminacyjnego od aktualnych, zdecentralizowanych form władztwa wpływa przy tym na przemianę wyartykułowanych w przepisach norm[17]. Z tej przyczyny badanie problemu dyskryminacji automatycznej uzasadnione jest również z perspektywy filozofii prawa, albowiem konieczna jest ponowna refleksja normatywna nad prawem antydyskryminacyjnym i funkcją aktualnych regulacji mających realizować postulaty równości leżącej u podstaw porządku prawnego nowożytnych demokracji. Uważam, że ta perspektywa może wzbogacać i uzupełniać badania doktrynalne w zakresie dyskryminacji automatycznej, a nadto zapewnić odpowiednią wnikliwość ocen.

W rozdziale I wykorzystuję metodę badawczą analizy autorskiej[18], przedstawiając teoretycznoprawne rozważania na temat dyskryminacji w ujęciu Freja Klema Thomsena, Tarunabha Khaitana, Kaspra Lipperta-Rasmussena czy Berta Heinrichsa z uwzględnieniem historycznego kontekstu pojęć dyskryminacji i równości. Poszukuję wyjaśnienia koncepcyjnego oraz pojęciowego, starając się przy tym zrekonstruować podstawowe cele zakazu dyskryminacji i dzięki temu zrozumieć przedmiot regulacji prawa antydyskryminacyjnego. Analiza ma prowadzić do odpowiedzi na pytanie, dlaczego pewne różnicowanie jest penalizowane społecznie i prawnie, podczas gdy inne takiemu nie podlega, co w konsekwencji prowadzi do ustalenia specyficznych cech dyskryminacji.

W rozdziale II przeprowadzam metaanalizę literatury poświęconej automatycznym systemom decyzyjnym, wprowadzając kontekst technologiczny do przedmiotu badania. Natomiast rozdział III ma na celu zaprezentowanie doktrynalnej rekonstrukcji obecnego systemu prawa antydyskryminacyjnego i regulacji mogących uzupełnić zakaz dyskryminacji w kontekście

[16] W. Moka-Mubelo S.J., *Towards a Contextual Understanding of Human Rights*, „Ethics & Global Politics" 12, 2019, nr 4, s. 40–52.
[17] J. Eckert, *What Is the Context in "Law in Context"?*, [w:] *Concepts of Law. Comparative, Jurisprudential, and Social Science Perspectives*, red. L.H. Urscheler, S.P. Donlan, London 2014, s. 1–12.
[18] Więcej zob. S. Taekema, W. van der Burg, *Legal Philosophy as an Enrichment of Doctrinal Research Part I: Introducing Three Philosophical Methods*, „Law and Method" 2020, s. 9.

9:81668

specyfiki technologicznej automatycznych systemów decyzyjnych. Podaję tu przykłady orzecznictwa związane z przedmiotem badania oraz zestawiam wnioski z teoretycznymi celami i założeniami regulacji.

Rozdział IV zawiera analizę doktrynalną projektu regulacji w sprawie sztucznej inteligencji z elementami analizy pojęciowej. Dochodzę w nim do wniosku, że niewytyczenie ram teoretycznych dla prawodawstwa regulującego cyfryzację może wpłynąć na upośledzenie oceny ryzyka, a nawet doprowadzić do błędnego doboru metod legislacyjnych[19]. Opisaną wcześniej metodę łączę tutaj z analizą opartą na równowadze refleksyjnej[20], rozważając konieczność zmiany techniki regulacji ze względu na specyficzne dla badanego problemu cechy automatycznych systemów decyzyjnych.

Z kolei w rozdziale V przeprowadzam teoretyczną analizę pojęciową z uwzględnieniem kontekstu historycznego i technologicznego. Badanie prowadzi mnie do krytycznej oceny instrumentów prawa, w związku z czym opisuję tu również swoje poszukiwania twórczych perspektyw i rekomendacji dla rozwoju teorii dyskryminacji automatycznej.

4. Struktura pracy

W rozdziale I przedstawiam historyczny kontekst wykształcania się zakresu podmiotowego i przedmiotowego prawa antydyskryminacyjnego w toku przełomowych walk o równość. Uzasadniam, że dotychczasowa logika zakazu dyskryminacji, definiująca jego treść katalogiem cech chronionych, doprowadziła do wielu problemów pojęciowych. Tę część podsumowuję próbą zrekonstruowania sposobu budowania definicji dyskryminacji i zaprezentowaniem jej form. Rozdział ten ma na celu ukazanie przede wszystkim perspektywy społeczno-analitycznej problemu dyskryminacji.

Uzasadnieniem osadzenia analizy na podłożu historycznym jest ukazanie odmienności problemów, na jakie prawo antydyskryminacyjne w społeczeństwie scentralizowanym miało stanowić odpowiedź. Walka o równość toczyła się bowiem na styku państwa i grup marginalizowanych. Procesem różnicowania dotknięte były głównie grupy historycznie słabsze pod względem ekonomicznym czy społecznym, jak również grupy niedostatecznie

[19] E.L. Sidorenko, P. von Arx, *Transformation of Law in the Context of Digitalization: Defining the Correct Priorities*, „Digital Law Journal" 1, 2020, nr 1, s. 24–38.
[20] Więcej zob. S. Taekema, W. van der Burg, *op. cit.*, s. 10.

10:62349

reprezentowane politycznie. Różnicowanie, któremu przeciwdziałać miało omawiane prawo, oparte było na widocznych cechach przyrodzonych człowieka, które następnie zyskały miano chronionych prawem.

Rozumienie pojęcia dyskryminacji wykształconego w toku walki o równość grup marginalizowanych prowadzi mnie do wniosku, że dominacja władzy publicznej nad społeczeństwem miała istotny wpływ na tworzenie i utrzymywanie stanów nierówności. Kontekst ten jest wyjątkowo ważny z perspektywy dalszej analizy i próby ukazania niekoherentności różnicowania algorytmicznego z formami, treścią i zakresem zakazu dyskryminacji objętego regulacjami prawa europejskiego. Na tym tle prezentuję definicję dyskryminacji z perspektywy społecznej oraz analitycznej, a także poszukuję logiki w wykształcaniu się cech chronionych prawem antydyskryminacyjnym. W ramach rozdziału I staram się przede wszystkim nakreślić fenomen polityczny, społeczny oraz prawny pojęcia dyskryminacji.

Celem rozdziału II jest przedstawienie ogólnych zasad działania automatycznych systemów decyzyjnych oraz wyzwań, z jakimi wiąże się ich stosowanie. W ramach analizy pochylam się nad wpływem różnicowania algorytmicznego na prawa i wolności jednostki oraz prezentuję definicje podstawowych pojęć, w tym: automatycznych systemów decyzyjnych, algorytmu, sztucznej inteligencji, uczenia maszynowego oraz dużych zbiorów danych (ang. *big data*). Automatyczny proces decyzyjny zestawiam tu z ludzkim procesem i ukazuję zarówno różnice, jak i podobieństwa obu mechanizmów. W ramach opisu zjawiska różnicowania algorytmicznego przybliżam zagrożenia społeczne związane ze skutkami marginalizacji grup historycznie wykluczonych oraz problem błędów algorytmicznych prowadzących do pogłębiania wykluczenia. Przedstawiając specyfikę zagrożeń, ukazuję jednocześnie masowość decyzji i ich odmienność związane z nieprzejrzystością algorytmów, niejawnością procesów decyzyjnych oraz ryzykiem różnicowania ze względu na zespoły cech, dane behawioralne lub kategorie nieuznawane za cechy chronione prawem antydyskryminacyjnym. Rozdział ten jest wstępem do dalszej analizy pojęć dyskryminacji automatycznej i wykluczenia algorytmicznego.

Rozdział III zawiera analizę regulacji prawnych, które na obecnym etapie mogą być źródłem zakazu różnicowania algorytmicznego. Zdając sobie sprawę z wielości globalnych reżimów regulujących ASD, koncentruję się na porządku europejskim. Wynika to z największego znaczenia tego porządku dla sytuacji polskiej, ale też intensywnych i zaawansowanych prac

11:35478

legislacyjnych trwających w Unii Europejskiej, wyznaczających globalnie kierunki myśli prawniczej wobec postępu technologicznego. Regulacje unijne będą kluczowe w najbliższej, dającej się przewidzieć przyszłości praw i obowiązków podmiotów stosujących algorytmy. Należy przy tym przypuszczać, że podobnie do rozporządzenia w sprawie ochrony danych osobowych (RODO) prawodawstwo UE będzie miało istotny wpływ na międzynarodowy rynek usług i produktów korzystających z systemów decyzyjnych, w tym algorytmów sztucznej inteligencji. W tym rozdziale ukazuję zatem drogę wykształcania się europejskiego prawa antydyskryminacyjnego, które rozwijało się partiami, w związku z czym ochrona antydyskryminacyjna poszerzała się, krocząc za zdarzeniami historycznymi związanymi z walką o równość.

W tej części analizuję również wydawane na tak wykształconym zakazie dyskryminacji orzecznictwo Trybunału Sprawiedliwości Unii Europejskiej oraz Europejskiego Trybunału Praw Człowieka pod kątem możliwości i zakresu zastosowania zakazu do problemu algorytmicznego różnicowania. Z uwagi na obserwowane ograniczenia prawa antydyskryminacyjnego uzupełniam analizę przepisami dotyczącymi ochrony danych osobowych i prawa do prywatności. W ramach rozdziału III poszukuję ponadto odpowiedzi na pytanie o treść zakazu dyskryminacji i zakres stosowania innych przepisów w kontekście ochrony jednostki przed różnicowaniem algorytmicznym.

Przedmiotem rozdziału IV są projekty aktów prawnych aktualnie procedowanych w Unii Europejskiej oraz instrumentów związanych z regulowaniem automatycznych systemów decyzyjnych. Pochylam się zatem nad drogą prowadzącą do powstania przepisów, a także zestawiam projektowane założenia z pytaniem o ich skutek dla różnicowania algorytmicznego i wykluczenia w specyficznej dla sztucznej inteligencji formule, stawiając jednocześnie pytania o adekwatność regulacji do masowości i nieprzejrzystości systemów. W tym kontekście przedstawiam logikę i stosowane techniki regulacyjne w projekcie rozporządzenia określanego jako Akt w sprawie sztucznej inteligencji. Egzemplifikuję odmienności w sposobie tworzenia prawa i wskazuję na dalsze możliwe kierunki rozwoju.

Równocześnie uwidoczniam konsekwencje istniejącej bariery epistemologicznej wynikającej z braku wglądu w sposób działania i ograniczenia w dekonstrukcji procesu decyzyjnego realizowanego przez zaawansowane algorytmy czarnoskrzynkowe. Stwierdzam, że bariera ta nie pozwala dokonywać racjonalnych wyborów na poziomie mechanizmów regulacyjnych. Dodatkowym problemem okazuje się nowość dziedziny i nieprzystawalność języka

12:27055

naturalnego do przedmiotu i adresata regulacji, co może utrudniać, a nawet uniemożliwiać oparcie się na klasycznych rozwiązaniach. Ukazuję w związku z tym powstawanie rozwiązań dotyczących ograniczenia negatywnego wpływu decyzji automatycznych na marginalizację i dyskryminację, biorąc pod uwagę, że był to rozwój zorientowany na przeciwdziałanie efektom istniejących klasycznie procesów dyskryminacji, a nie mechanizmów dyskryminacji automatycznej. Jednocześnie wskazuję na nieistnienie odpowiedniej regulacji szczególnej w projektowanych aktach, obejmującej problem specyfiki dyskryminacji automatycznej. Rozdział IV zawiera zatem krytyczną ocenę projektowanych przepisów ze względu na to, że ich założenia nie korespondują dostatecznie ze specyfiką dyskryminującego różnicowania automatycznego.

W rozdziale V staram się natomiast znaleźć odpowiedź na pytania o dalsze możliwe kierunki przeciwdziałania dyskryminacji w systemach automatycznych. Zestawiam tu wyróżniki europejskiego zakazu dyskryminacji z dyskryminacją automatyczną i na tym gruncie podejmuję próbę skonstruowania definicji tego typu dyskryminacji. Jednocześnie wobec intersekcyjności i emergentności różnicowania algorytmicznego, a także szerokiej delegacji uprawnień publicznoprawnych do stosowania automatycznych systemów decyzyjnych wskazuję, że ryzyko nowego wymiaru dyskryminacji dla społeczeństwa jest trwałe i inherentne. Stwierdzam, iż nie może być ono ograniczane konwencjonalnymi formami wykonywania władzy publicznej i nadzoru, w związku z czym konieczna jest nie tylko refleksja nad sposobem uzupełnienia materialnoprawnej treści zakazu dyskryminacji, ale również nałożenie obowiązków z zakresu autolustracji na podmioty stosujące ASD. Sugeruję ponadto poszukiwanie narzędzi i schematów umożliwiających pełniejszą kontrolę i egzekwowanie zakazu dyskryminacji automatycznej. Ograniczenie działania do projektowania regulacji ryzyka według mnie nie jest bowiem wystarczającą odpowiedzią na problem dyskryminacji automatycznej.

Zawarte w tej części pracy refleksje prowadzą mnie do uznania konieczności wypełnienia luki regulacyjnej poprzez rekonstruowanie roli i sposobu realizacji postulatu równościowego w ramach odpowiedzi na zagrożenia związane z ASD. Automatyczne systemy decyzyjne mogą być wszak istotnym elementem do głębszego zrozumienia problemu dyskryminacji i jego pierwotnej osi związanej z marginalizacją oraz wykluczeniem grup historycznie pokrzywdzonych i niedoreprezentowanych politycznie.

13:25713

I. Różnicowanie a dyskryminacja. Historyczne ujęcie problemu

Współczesne państwa demokratyczne zaliczają równość do swoich podstawowych zasad ustrojowych, nadając jej wagę fundamentu demokratycznego porządku prawnego. Idea równości, mimo wykształcenia się w czasach starożytnych, nie zawsze jednak prowadziła do symetrycznych relacji.

Pierwsze rozwinięte koncepcje równości formułowali Grecy (Arystoteles wskazywał, że równość oznacza równe traktowanie równych, a nierówne nierównych) i Rzymianie, przy czym znaczenie równości miało ówcześnie jedynie sens etyczny i nie prowadziło do właściwego równouprawnienia w zakresie przyznanych praw i wolności[21]. Aktualny kształt i treść równości ukształtowały natomiast koncepcje prawnonaturalne, pojmujące równość jako element godności człowieka, a zatem jedną z przyrodzonych cech każdej istoty ludzkiej[22]. Istotą zasady równości jest spowodowanie tożsamej sytuacji adresatów norm prawnych w procesach stanowienia i stosowania prawa[23].

Natomiast dyskryminacja[24] jest powszechnie rozumiana jako naruszenie tej zasady na skutek niekorzystnego różnicowania grupy osób względem innych, znajdujących się w podobnej lub tożsamej sytuacji faktycznej[25]. Treść zasady równości wykracza poza zakaz dyskryminacji, zawiera bowiem w sobie nie tylko obowiązek równego traktowania, czyli ochrony przed dyskryminacją, ale również pozytywne działania prowadzące do aktywnego przeciwdziałania nierówności[26]. W tym znaczeniu zasada niedyskryminacji, jako jedna z form dążenia do materializacji idei równości, jest filarem państwa prawa, umożliwiającym realizację postulatów sprawiedliwości społecznej[27].

Tak rozumiana sprawiedliwość nie oznacza jednak równości totalnej między jednostkami. Ronald Dworkin, pisząc o równości, wskazywał, że na równość składają się prawo do równego

---

[21] M. Derlatka *et al.*, *Konstytucja Rzeczypospolitej Polskiej. Komentarz*, red. L. Garlicki, M. Zubik, t. 2, Warszawa 2016, art. 32.
[22] *Ibidem*.
[23] K. Michałowska, *Różnicowanie a dyskryminacja – status prawny osób szczepionych i nieszczepionych na COVID-19*, „Przegląd Sądowy" 4, 2021, s. 38–56.
[24] Słowo „dyskryminować" pochodzi od łacińskiego czasownika *discriminare*, które z kolei wywodzi się od rzeczownika *discrimen*, oznaczającego 'rozróżnienie, różnicę, oddzielanie oraz porządkowanie'. Tym samym *discriminare* pierwotnie znaczyło 'rozróżniać, różnicować, oddzielać czy wyodrębniać'.
[25] D. Morondo Taramundi, *Discrimination by Machine-Based Decisions: Inputs and Limits of Anti-Discrimination Law*, [w:] *Law and Artificial Intelligence: Regulating AI and Applying AI in Legal Practice*, red. B. Custers, E. Fosch-Villaronga, The Hague 2022, s. 76.
[26] A. Śledzińska-Simon, *Zasada równości i zasada niedyskryminacji w prawie Unii Europejskiej*, „Studia BAS" 2, 2011, nr 26, s. 43.
[27] *Ibidem*.

14:49453

traktowania, czyli równego rozłożenia szans, zasobów czy obciążeń, oraz prawo do traktowania jako równego innym, tj. poszanowanie równe temu, jakie okazuje się każdemu innemu człowiekowi. Autor zauważał przy tym, że wykluczenie i różnicowanie jest immanentnym elementem prawa jako systemu[28]. Prawo nie przyznaje bowiem takich samych uprawnień wszystkim w każdej sytuacji[29], legalizując ze względu na ważne interesy ogółu lub jednostki pozbawienie lub ograniczenie poszczególnych praw.

Dyskryminacja to ocena prawna zjawiska różnicowania istniejącego mimo postulatów równościowych w kontekście negatywnego wpływu tego różnicowania na strukturę społeczną i demokrację. W takim znaczeniu różnicowanie będzie sprzeczne z prawem, gdy bezpośrednio lub pośrednio doprowadzi do nieakceptowanego i społecznie szkodliwego pokrzywdzenia członków grup społecznych ze względu na posiadanie przez nich konkretnych cech, takich jak płeć, rasa lub pochodzenie etniczne, religia, orientacja seksualna czy niepełnosprawność. Tak ustalone cechy dotykają bowiem bezpośrednio sfery godności istoty ludzkiej i z tej przyczyny nie mogą być podstawą kategoryzowania ludzi w dostępie do praw lub możliwości. Ewentualne uzasadnienie odmiennego traktowania podmiotów zidentyfikowanych pośrednio lub bezpośrednio na podstawie takich cech chronionych wymaga wskazania konstytucyjnego uzasadnienia. Różnicowanie powinno przy tym znajdować oparcie w jasno wyodrębnionych, a także relewantnych kryteriach, takich jak racjonalność, proporcjonalność oraz sprawiedliwość dokonywania zróżnicowań[30].

Należy też wskazać, że dyskryminacja nie jest zjawiskiem statycznym i podlega ciągłej ewolucji wraz z systemami wartości leżącymi u podstaw danego prawodawstwa, zmieniając swój wymiar i treść w toku urzeczywistniania postulatów równościowych, implementowanych w społeczeństwach demokratycznych. Kształt zakazu dyskryminacji, jako jeden z przejawów emanacji zasady równości, formułował się zatem wraz ze zmianą pojmowania roli jednostki w życiu społecznym, gospodarczym i politycznym z powodu odrzucenia dotychczasowej organizacji społeczeństwa.

Za Zygmuntem Baumanem[31] można przyjąć, że natura niedoli, którą cierpiały grupy marginalizowane, prowadziła do otwartego sprzeciwu. Walka o równość przenikała więc z pierwotnej relacji istniejącej na styku pokrzywdzonych z władzą na pozostałe grupy społeczne,

[28] R. Dworkin, *Biorąc prawa poważnie*, przeł. J. Woleński, Warszawa 1998, s. 407.
[29] W prawie występuje np. ograniczenie praw wyborczych osób niepełnoletnich, ograniczenie wieku przy nabywaniu alkoholu lub oglądaniu treści zawierających przemoc.
[30] Wyrok Trybunału Konstytucyjnego (dalej: TK) z 16.12.1997 r., sygn. K 8/97, OTK 1997, Nr 5–6, poz. 70.
[31] Z. Bauman, *Retrotopia. Jak rządzi nami przeszłość?*, Warszawa 2018, s. 163.

15:11368

uruchamiając procesy uświadamiania niesprawiedliwości związanej z marginalizowaniem i wykluczeniem poszczególnych grup. Krzywda grup walczących o równość była widoczna na tle społecznym, a niezgoda na utrwalone historycznie procesy marginalizacji poruszała nie tylko ludzi identyfikujących się z danymi grupami.

Dlatego też twierdzę, że cel i funkcja zakazu dyskryminacji zostały ustalone na podłożu politycznym i społeczno-kulturowym, w związku z czym odpowiadają one historycznie wykształconemu zakresowi prawa do niepodlegania krzywdzącemu różnicowaniu. W tym sensie przeciwdziałanie dyskryminacji ma formę wieloetapową, a zakaz dyskryminacji można uznać za element kultury, który wraz z nią podlega stałym zmianom.

W niniejszym rozdziale prezentuję eksplikację logiki dyskryminacji i wiążącą się z nią historię walki o równość, a na tym tle dyskusje co do kształtu, form i treści zakazu dyskryminacji.

1. Historyczna walka o równość

Źródłem dyskryminacji są relacje społeczne wykształcone na podstawie istniejących stosunków władzy i podległości, które cechowała społeczna niesprawiedliwość. W tym kontekście związek prawa antydyskryminacyjnego z historią współczesnych demokracji liberalnych jest ścisły. Demokracje liberalne z jednej strony zakładają bowiem możliwość różnic w zakresie rasy, religii czy preferencji seksualnych, a z drugiej uznają, że rolą władz publicznych jest traktowanie wszystkich obywateli równo. Jednocześnie każda władza polityczna posługuje się licznymi administracyjnymi i publicznymi kategoriami oraz hierarchiami różnicującymi, co zwiększa jej efektywność. Rolą prawa antydyskryminacyjnego było więc wyznaczenie granic dozwolonego pozytywnego i negatywnego nierównego traktowania. Z tej perspektywy zakaz dyskryminacji był integralną częścią zasady równości. W celu realizacji możliwie szerokich postulatów równościowych na gruncie zakazu dyskryminacji sformułowano katalogi cech chronionych oraz przesłanek, których spełnienie warunkuje możliwość uzasadnienia odmiennego traktowania adresatów decyzji. Zakaz różnicowania ze względu na określone cechy wykształcał się przy tym stopniowo i kontekstowo w toku serii konfliktów politycznych i walk o prawa podstawowe, będących przejawem postępującej równości i wolności.

Jedną z podstawowych cech chronionych zakazem dyskryminacji jest rasa człowieka. Punktem wyjścia do dalszej analizy będą w związku z tym zdarzenia poprzedzające amerykańskie

16:39552

manifesty praw obywatelskich, jako że deklaracje amerykańskie stały się prologiem późniejszego urzeczywistniania idei demokracji w Europie, nadając jej kierunek i kształt.

W przypadku USA etnocentryzm, wyrażający się w idealizowaniu wzorców społecznych panujących w imperiach kolonialnych, miał swoje początki w hierarchicznych relacjach między białymi osadnikami Wirginii a ludnością afroamerykańską sprowadzaną do USA w roli niewolników. Pozyskiwanie i przesiedlanie mieszkańców Afryki do pracy niewolniczej było wówczas przyjmowane jako naturalny i społecznie uzasadniony proces umożliwiający utrzymanie obowiązujących w tamtym okresie struktur i relacji społeczno-majątkowych. Na skutek tego, mimo początkowego niezalegalizowania niewolnictwa, u białej ludności ugruntowało się szczególne poczucie władztwa i dominacji związane z zależnością grup podporządkowanych – łączące nienawiść, wzgardę, litość i protekcjonalność wobec Afroamerykanów[32]. Nastawienie to potęgował kreacjonizm, który nadawał istniejącemu porządkowi społecznemu miano niepodważalnej boskiej koncepcji, co jedynie wspierało teorie rasistowskie. Egzekwowanie władztwa w stosunku do niewolników odbywało się przy wykorzystaniu przemocy i brutalnych praktyk, mających utrwalać istniejący porządek zwłaszcza poprzez uprzedmiotowienie i absolutne podporządkowanie zdominowanych społeczności.

Niewolnictwo jako filar systemu plantacji umacniało się wraz z jego rozkwitem[33]. Wynikało to z jednej strony z niedoborów siły roboczej na terenie USA, a z drugiej – z perspektywy ekonomicznej, tj. braku obciążeń socjalnych i finansowych wobec zniewolonej ludności. Niewolnictwo stało się zatem mechanizmem niezwykle opłacalnym dla białych, z którego mimo jawnej niesprawiedliwości społecznej trudno było zrezygnować (nawet na poziomie państwowym). Istniejącym praktykom towarzyszyła jednak nasilająca się obawa buntu i sprzeciwu niewolników wobec niesprawiedliwości i wyzysku. Z tej przyczyny początkowo starano się zachować istniejące stosunki dominacji przez wpisanie jej w prawo, np.

[32] Prawo uchwalone w 1639 r. stanowiło, że „wszyscy z wyjątkiem Afroamerykanów" mają otrzymać broń i amunicję (prawdopodobnie do walki z rdzennymi mieszkańcami Ameryki). Gdy w 1640 r. trzech pracowników służebnych próbowało ucieczki, karą dla dwóch białych mężczyzn było przedłużenie ich służby, natomiast – jak czytamy w orzeczeniu sądu – „trzeci z nich, czarny nazwiskiem John Punch, będzie służyć swemu panu lub osobie przez niego wyznaczonej przez całe swe życie". W tym samym roku odnotowano przypadek czarnoskórej służącej, której dziecko spłodził biały mężczyzna, niejaki Robert Sweat. Sąd orzekł, że kobietę należy „wychłostać pod pręgierzem, zaś Sweat powinien nazajutrz przed południem pokajać się publicznie za swój występek w kościele miejskim James[town]" – H. Zinn, *Ludowa historia Stanów Zjednoczonych. Od roku 1492 do dziś*, przeł. A. Wojtasik, Warszawa 2016, s. 73.

[33] W Wirginii w 1763 r. liczba niewolników wzrosła do 170 tys. z początkowych 6 tys. w 1700 r.

17:39528

zalegalizowanie fizycznego karania niewolników za nieposłuszeństwo[34]. Tymczasem hermetyczność kręgu najbogatszych, wspomagana drakońskimi przywilejami grupy dominującej, wpływała na dalej idącą nierówną dystrybucję dóbr materialnych i bogactw. Niewolnictwo w dużej części oparto w końcu na aparacie prawnym o charakterze globalnym.

Rasizm wzmacniany był przy tym klasowym strachem[35]. Przedstawiciele władz obawiali się bowiem, że w obliczu skali cierpienia sprzeciw okaże się nieunikniony. Rozkład społeczny i wzbierająca fala okrucieństwa doprowadziły wreszcie do wybuchu aktów niezgody, przybierających formę zamieszek. Widocznych reakcji społecznych nie mógł już powstrzymać strach przed stosowaniem drakońskiego prawa i kar fizycznych.

Na tym tle od czasów rebelii Bacona w Wirginii do lat 60. XVIII w. wybuchło 18 powstań mających na celu obalenie istniejącej władzy. Działania rewolucyjne dotknęły wszystkich beneficjentów kolonializmu, jako że w miastach kolonialnych wykwalifikowani robotnicy postulowali ustanowienie demokracji politycznej, a więc otwartych posiedzeń zgromadzeń przedstawicielskich, galerii dla publiczności w salach sądowych, a także ujawniania przebiegu głosowań imiennych, tak aby wyborcy mogli kontrolować swoich reprezentantów[36].

W tych okolicznościach jasne stało się, że istniejący system upada, przez co nieodzowna jest reakcja prawa na eksplodujące problemy społeczne, w tym właśnie kwestię rasizmu[37]. Odpowiedzią miały być Deklaracja praw Wirginii z 12.06.1776 r. oraz Deklaracja niepodległości z 4.07.1776 r. Wskazane akty prawa gwarantowały ochronę praw białych mężczyzn, a zatem nawet nie zbliżały się do skali realnego problemu niewolnictwa i rasizmu.

---

[34] W prowincji Maryland w 1723 r. przyjęto regulację, na mocy której niewolnikowi obcinano uszy w wypadku nieposłuszeństwa, poważniejsze przestępstwa zaś mogły zostać ukarane przez powieszenie, ćwiartowanie i wystawienie zwłok na widok publiczny; zob. *ibidem,* s. 79.

[35] Około 1700 r. Izba Obywatelskiej Wirginii oznajmiła: „Chrześcijańscy pracownicy służebni w tym kraju w przeważającej części składają się z gorszego rodzaju Europejczyków. A fakt, że […] bardzo wielu ze sprowadzonych Irlandczyków i przedstawicieli innych nacji uczestniczyło w niedawnych wojnach jako żołnierze, w naszych obecnych warunkach niemal uniemożliwia nam rządzenie nimi, gdyby zaś zostali oni uzbrojeni i mieli okazję zebrać się razem, możemy mieć uzasadnione obawy, że byliby gotowi wystąpić przeciwko nam" – *ibidem,* s. 122 i n.

[36] *Ibidem,* s. 117.

[37] Rasizm (franc. *La race* – ród, grupa spokrewniona) – zespół poglądów, według których naturalne i wrodzone różnice między rasami ludzkimi prowadzą do dominacji jednej rasy nad pozostałymi. Obecnie uznawany jest za jedną z doktryn politycznych i społecznych, która – choć nie znajduje uzasadnienia w naukach przyrodniczych czy społecznych – zdobyła znaczącą popularność w XIX i XX w. Pojęcie „dyskryminacja rasowa" oznacza wszelkie działania polegające na zróżnicowaniu mającym formę wykluczenia bądź uprzywilejowania z powodu rasy, koloru skóry, urodzenia, pochodzenia narodowego lub etnicznego, które ma na celu lub pociąga za sobą naruszenie zasady równości z katalogu praw człowieka i podstawowych wolności w sferze politycznej, gospodarczej, społecznej i kulturalnej lub w innych dziedzinach życia prywatnego lub publicznego; zob. Międzynarodowa konwencja w sprawie likwidacji wszelkich form dyskryminacji rasowej z 6.09.1969 r., Dz.U. Nr 25, poz. 187, art. I, par. 1.

18:10672

Niepokój społeczny świadczył natomiast o dezaktualizowaniu się schematów leżących u podstaw ówczesnego systemu władztwa, co uzasadniało konieczność szerszych zmian. Jedynie odgórne ukształtowanie nowych zasad dawało nadzieję na uniknięcie krwawej rewolucji, zagrażającej ówczesnym klasom rządzącym. Działania podjęte w wyniku społecznego sprzeciwu sięgnęły jedynie do poziomu, w którym zmiany nie naruszały innych istotnych interesów władzy. Co prawda, zniesienie niewolnictwa było momentem przełomowym i rewolucyjnym, jednak nie sposób utożsamiać tego z zamknięciem problemu uprzedzeń, marginalizacji czy segregacji rasowej w USA, którego emanacje widoczne były w ruchu praw obywatelskich w XX w., a nawet są dostrzegalne obecnie.

Również w Europie porządek prawny przez stulecia determinował stan nierówności społecznej. Prawa i obowiązki jednostki wyznaczane były bowiem na podstawie jej przynależności do konkretnej grupy społecznej (np. pochodzenia, stanu, kręgu wolno urodzonych)[38], podczas gdy pozycja danej osoby była określana w odniesieniu do takich aspektów jak kapitał ekonomiczny, społeczny czy też kulturowy jednostki lub poprzez wartościowanie okoliczności związanych z jej miejscem urodzenia lub pochodzeniem. Na gruncie utrwalonego w tamtym okresie elitaryzmu[39] dochodziło w istocie do wykluczenia społecznego grup niezamożnych lub niezwiązanych z kręgami władzy, co w dalszej perspektywie spowodowało walki o równość na tle pochodzenia społecznego[40].

Przełom zapoczątkowało rozwinięcie koncepcji praw podmiotowych jednostki w dobie rewolucji francuskiej. Droga ta zakładała przemianę jednostki w obywatela, którego prawa podmiotowe były sukcesywnie uznawane przez władzę publiczną tak, że jednostka na tle pozostałych współobywateli zyskiwała prawo do równości wobec prawa. Taka zmiana była zaprzeczeniem treści realiów stanowych i wynikających z nich ówcześnie instytucji obywatelstwa miejskiego[41].

Centralnym powodem zmian z 1789 r. były skostniałe formy władzy absolutnej, która blokowała autonomię poddanych i kiełkujące zmiany społeczne w kierunku świadomości

[38] K. Trzciński, *Obywatelstwo w Europie. Idea i jej wyraz formalny w perspektywie historycznej,* „Studia Europejskie" 2, 2002, s. 46–55.
[39] Elitaryzm (franc. *élire* < łac. *eligere* – wybierać) – wyodrębnienie w społeczeństwie grup kwalifikowanych jako elity i nadanie im uprzywilejowanego znaczenia względem pozostałych obywateli.
[40] Dyskryminacja z uwagi na pochodzenie społeczne oznacza, że osoba jest różnicowana ze względu na swoją klasę, kastę lub kategorię społeczno-zawodową. Nie istnieje jednak ujednolicona definicja klasy czy pochodzenia społecznego; w tym zakresie funkcjonuje obok siebie wiele konkurujących z sobą teorii (najbardziej rozpowszechnione są koncepcje Karola Marksa, Maxa Webera, Emile'a Durkheima i Pierre'a Bourdieu).
[41] K. Trzciński, *Początki nowożytnego obywatelstwa w Europie – obywatel państwa i katalog jego praw w dokumentach Rewolucji Francuskiej,* „Studia Europejskie" 2, 2005, s. 92.

19:83582

obywatelskiej. Przyrost problemów wewnętrznych związanych z kryzysem ekonomicznym i rolnym wyzwolił we Francji rozgrzane nastroje społeczne.

Rewolucyjna konstytuanta 26.08.1789 r. przyjęła Deklarację praw człowieka i obywatela. Dokument ten w historii prawodawstwa europejskiego miał unikatowy charakter, albowiem stał się zaczynem nowożytnej idei demokracji. Wzorców francuskiej deklaracji należy jednak upatrywać przede wszystkim w amerykańskich manifestach praw obywatelskich[42].

Deklaracja praw miała zharmonizować prawa naturalne i prawa pozytywne oraz być kompromisem między tym, co ludzkie i co obywatelskie. Relewantna dla przedmiotu niniejszej analizy równość wobec prawa została wyrażona już w pierwszym artykule deklaracji. Równość wolnych ludzi postrzegana była ówcześnie jako źródło powszechnych praw politycznych obywateli. Oznaczało to możliwość osobistego lub pośredniego udziału w procesie legislacyjnym, a także dostęp do stanowisk i urzędów publicznych, ograniczony jedynie osobistym uzdolnieniem oraz zasługami jednostki.

Akt ten zapoczątkował jednocześnie gruntowną reformę na poziomie zarządzania terytorialnego, stając się pierwszą fazą realnej decentralizacji władzy[43]. Mimo przełomowego poszerzenia wolności obywatelskich zasada równości formalnie była jednak korygowana statusem majątkowym i wysokością dochodów jednostki. Status obywateli czynnych otrzymało bowiem jedynie ok. 4,3 mln osób, a zatem losy państwa nadal pozostawały w rękach uprzywilejowanych grup społecznych[44]. Sprzeciw obywateli biernych wobec ich pominięcia w dokumencie znalazł w końcu wyraz w powstaniu ludowym rozpoczętym 10.08.1792 r. W jego efekcie poszerzono grupy obywateli uprawnione do głosowania poprzez przyznanie praw praktycznie każdemu mężczyźnie, który ukończył 21 lat. W kręgu obywateli biernych nadal pozostali kobiety i nieletni.

Rewolucja francuska nie tylko zatrzymała się na kryteriach materialnych, ale też praktycznie żadne przeobrażenia nie nastąpiły na płaszczyźnie praw politycznych kobiet. Pozornie były one równe w prawach z mężczyznami, lecz w treści aktów całkowicie je pominięto. Dało to asumpt Olimpii de Gouges do ogłoszenia we wrześniu 1791 r. Deklaracji praw kobiety i obywatelki[45]. Manifest kierowany był wyłącznie do kobiet, a w treści wyrażał niezadowolenie

[42] J. Baszkiewicz, *1789: Spory o prawa człowieka i obywatela*, [w:] *Człowiek jako obywatel*, red. M. Szyszkowska, Warszawa 1995, s. 207–208.
[43] K. Trzciński, *Początki nowożytnego obywatelstwa*…, s. 71.
[44] K. Koranyi, *Powszechna historia państwa i prawa*, t. 4, Warszawa 1967, s. 251.
[45] O. de Gouges, *Déclaration Des Droits de La Femme et de La Citoyenne*, [b.m.w.] 1791.

20:74600

z prawnie usankcjonowanej społecznej dominacji mężczyzn. Krytyka wyrażana przez de Gouges i jej jasno formułowany sprzeciw wobec męskiej hegemonii były bezpośrednią przyczyną orzeczenia i wykonania na niej kary śmierci[46]. Doświadczenia rewolucji uwidoczniły tymczasem potrzebę gruntownej zmiany postrzegania roli kobiet w państwie, nie tylko na poziomie społecznym, ale również politycznym. Niedoreprezentowanie kobiet w strukturach władzy uczyniło jednak ich postulaty niesłyszalnymi przez mężczyzn w toku rewolucji.

Walka o równość płci na tle szerzącego się zjawiska seksizmu[47] i społecznego wykluczenia kobiet były przy tym tłem wykształcania się pojęcia dyskryminacji z uwagi na płeć. Jako że prawo stanowiło w tym zakresie swoistą podstawę nierówności, w XVIII w. ruch emancypacji kobiet otwarcie sprzeciwił się istniejącym schematom społecznym. Kobiety domagały się statusu obywatelskiego równego pozycji mężczyzn w społeczeństwie[48].

W 1782 r. głos w dyskusji zabrała Mary Wollstonecraft[49], która w fundamentalnym dla francuskiego i brytyjskiego feminizmu manifeście przedstawiła propozycje polityczne i edukacyjne. Zmiany te miały odmienić istniejący porządek społeczny. Postulowane transformacje nie zostały przeprowadzone. Co więcej, kodeks Napoleona umocnił marginalizowanie kobiet, przyznając ich mężom prawa do majątku i dzieci, a w razie zdrady sankcjonując stosowanie wobec małżonek kar fizycznych. Równocześnie praca kobiet traktowana była jako niepełnowartościowa, czym uzasadniano zaniżanie ich wynagrodzeń.

Zniewolenie kobiet wynikające z osadzenia seksizmu w instytucjach społecznych, męskiej dominacji w relacjach rodzinnych i utrudnień w uzyskaniu niezależności finansowej przez przedstawicielki tej płci istotnie utrudniało im drogę do równości. Feministki walczące o prawa i wolności dostrzegły analogię między sytuacją niewolników w państwach kolonialnych a własnym seksualnym, prawnym, emocjonalnym, a także fizycznym zniewoleniem przez mężczyzn[50]. Żądania głoszone w ramach programów emancypacji kobiet obejmowały w związku z tym walkę z innymi formami wykluczenia, w tym postulaty odnoszące się do kwestii

[46] K. Trzciński, *Początki nowożytnego obywatelstwa*…, s. 86.
[47] Seksizm opiera się na założeniu, że dominacja mężczyzn nad kobietami jest obiektywnie uzasadniona przymiotami i cechami, którymi dysponują mężczyźni. Termin został stworzony przez feministki w celu ukazania podobieństwa między rasizmem a negatywnymi stereotypami dotyczącymi schematów społecznych stosowanych wobec kobiet. Seksistowskie postawy prowadzą bowiem do dyskryminacji, jak również takich zachowań jak molestowanie seksualne, molestowanie uliczne i nieuczciwe praktyki zatrudniania; zob. J.K. Boles, D.L. Hoeveler, *Historical Dictionary of Feminism*, Lanham 2004, s. 291.
[48] K. Trzciński, *Początki nowożytnego obywatelstwa*…, s. 92.
[49] M. Wollstonecraft, *A Vindication of the Rights of Woman*, Boston 1792.
[50] J. Hannam, *Feminizm*, przeł. A. Kaflińska, Poznań 2010, s. 38.

21:33389

przyznania praw robotnikom, zniesienia niewolnictwa, ustanowienia reprezentatywnych rządów czy reformy majątkowej. Akcjom feministycznym towarzyszyły zatem procesy związane z ogólnym rozwojem społeczno-gospodarczym, wynikającym przede wszystkim ze zwiększenia możliwości i dostępu kobiet do edukacji.

Ich niezgoda zmieniła charakter na bardziej radykalny, gdy dwie amerykańskie feministki zostały aresztowane za skandowanie hasła „Głos dla kobiet!" podczas kampanii wyborczej w 1905 r. Na skutek tych działań nagromadzona frustracja przerodziła się w rewolucyjną kampanię, podczas której dochodziło do wybijania okien czy podpalania opuszczonych budynków. Pasja i odwaga sufrażystek stały się impulsem dla wielu kobiet, by całkowicie oddać się walce o prawa wyborcze i swobody obywatelskie. Bunt rozlał się na wszystkie stany pochodzenia.

Ostatecznie uzyskanie praw wyborczych było pewnym przełomem, który nie wpłynął jednak zasadniczo na sytuację faktyczną kobiet. Simone de Beauvoir twierdziła, że przypisywane przedstawicielkom płci żeńskiej role i cechy zostały utworzone społecznie[51], co w konsekwencji uniemożliwiło wyczerpanie postulatów feministek w okresie powojennym. Podobnie jak w przypadku niewolnictwa walka o równość kobiet zatrzymała się tam, gdzie postawiono granicę nienaruszalności interesów patriarchalnego społeczeństwa. Co więcej, odpowiedź prawa ograniczyła się do regulacji i zabezpieczenia interesów białych kobiet pochodzących z klasy średniej. Pominięto przy tym inne marginalizowane grupy i nakładające się na siebie relacje krzyżowe wpływające na zwielokrotnienie niesprawiedliwości, z jaką mierzyły się kobiety zaliczane jednocześnie do innych marginalizowanych klas społecznych czy grup etnicznych. Walka o równość kobiet nadal trwa, a deficyty poznawcze i społeczne związane z marginalizacją tej grupy społecznej mają swoje konsekwencje do dziś[52].

Kolejnym niezmiernie istotnym elementem walki o równość był sprzeciw wobec wykluczenia społeczności nieheteroseksualnej z życia publicznego. Wobec uznania przedstawicieli społeczności LGBT za jednostki zagrażające bezpieczeństwu Ameryki grupa ta była zmuszona podjąć walkę o równość. W latach 60. i 70. XX w. zrodziło się pojęcie homofobii, rozumianej jako nieuzasadniona niekorzystna postawa wobec zachowań nieheteroseksualnych. Niestety początkowo problematyka homofobii[53] była tematem tabu na arenie międzynarodowej.

[51] S. de Beauvoir, *Druga płeć*, t. 2, przeł. M. Leśniewska, Kraków 1972, s. 11.
[52] C. Criado Perez, *Niewidzialne kobiety*, przeł. A. Sak, Kraków 2020.
[53] Termin „homofobia" rozumiany jest jako negatywna ocena osób homoseksualnych, grup społecznych i wszelkich zachowań czy atrybutów kulturowo z nimi kojarzonych. Na poziomie społecznym wiąże się ona z negatywną oceną i założeniem niższego statusu nieheteroseksualnych zachowań, tożsamości, relacji czy

22:89699

Homoseksualiści byli częściowo traktowani jak jednostki chore i polegające obligatoryjnemu leczeniu. Metodami stosowanymi w walce z niechcianymi zachowaniami seksualnymi było przymusowe leczenie psychiatryczne, terapie hormonalne, a nawet kastracja. Takie terapie były stosowane w USA aż do 1973 r., kiedy to Amerykańskie Towarzystwo Psychiatryczne odstąpiło od kwalifikowania skłonności homoseksualnych jako zaburzenia[54].

Niestety zaprzestanie aplikowania praktyk leczniczych wobec członków społeczności LGBT nie oznaczało w żadnym wymiarze uznania równości tych osób wobec prawa. Kluczowy okazał się czerwiec 1969 r., gdy geje przeciwstawili się fizycznie przygotowanemu przez władzę nalotowi policji, która wtargnęła w trakcie imprezy do klubu Stonewall Inn w dzielnicy Greenwich Village na nowojorskim Manhattanie. Na skutek sprzeciwu i protestów doszło wówczas do trzydniowych zamieszek z udziałem przedstawicieli LGBT i policji. Powstanie w Stonewall Inn uważa się obecnie za formalny początek rewolucji i wykształcenie się ruchu obrony praw LGBT. Społeczność homoseksualna, podobnie jak 14 lat wcześniej Afroamerykanie[55], rozpoczęła wówczas walkę o poszanowanie praw człowieka, a sprzeciw ten – jak w innych opisywanych przypadkach – wynikał z widocznej i publicznej marginalizacji, pokrzywdzenia i wykluczenia przedstawicieli LGBT w relacjach z władzą publiczną.

Wydarzenia w Stonewall Inn poskutkowały powołaniem do życia w lipcu 1969 r. Gay Liberation Front, czyli Gejowskiego Frontu Wyzwolenia, pełniącego funkcję centrum walki o równouprawnienie. Uznaje się przy tym, że bez zdarzeń towarzyszących tzw. Stonewall nie wykreślono by w 1973 r. homoseksualizmu z listy chorób ani nie zalegalizowano w 2015 r. małżeństw jednopłciowych w niektórych stanach USA. Nierówności i marginalizacja nie zakończyły się jednak na konsekwencjach omawianych wydarzeń. Społeczność LGBT aktualnie również mierzy się z licznymi uprzedzeniami i stygmatami.

Analizując te przykładowe historyczne zdarzenia, które wpłynęły na nowożytne rozumienie zakazu dyskryminacji, należy wskazać, że wykształcenie się cech chronionych, np. rasy, pochodzenia społecznego, płci czy orientacji seksualnej, nastąpiło w procesie różnicowania ze względu na obserwowalne cechy ludzkie. Jego efektem było jawne i publiczne ograniczenie

zbiorowości; zob. G.M. Herek, *Beyond "Homophobia": Thinking about Sexual Prejudice and Stigma in the Twenty-First Century*, „Sexuality Research & Social Policy" 1, 2004, s. 6–24; *idem*, S*cience, Public Policy, and Legal Recognition of Same-Sex Relationships*, „American Psychologist" 62, 2007, nr 7, s. 713–715.

[54] Proces ten obrazuje DSM (*The Diagnostic and Statistical Manual of Mental Disorders*), zmieniająca się w czasie klasyfikacja zaburzeń psychicznych, publikowana przez Amerykańskie Towarzystwo Psychiatryczne.

[55] Rosa Parks – czarnoskóra szwaczka – 1.12.1955 r. wbrew ówcześnie obowiązującemu prawu odmówiła ustąpienia miejsca białej osobie w miejskim autobusie. Zdarzenie to zapoczątkowało masowe protesty, którym przewodniczył Martin Luther King Jr.

23:89549

pewnych kategorii osób w dostępie do praw i dóbr społecznych, takich jak uprawnienie do dokonywania określonych czynności prawnych, możliwość egzekucji praw czy korzystania z przestrzeni publicznej.

Równość zawsze miała charakter akcesoryjny, co oznacza, że nie istniała w oderwaniu od konkretnego uprawnienia (prawa polityczne, prawa pracownicze, własność, ochrona zdrowia). Same procesy różnicowania były jednak widoczne zarówno dla grupy uprzywilejowanej, jak i pokrzywdzonej i odbywały się na podstawie prawa stanowionego przez scentralizowaną władzę publiczną. Aparat prawny odgrywał przez to konstytutywną rolę w dokonywaniu i podtrzymywaniu różnicowania.

Niesprawiedliwość różnicowania była przedmiotem sprzeciwu wyrażanego jawnie i publicznie przez podmiot zbiorowy, realizujący swoje postulaty równościowe w toku działań rewolucyjnych. Adresatem sprzeciwu była przede wszystkim władza publiczna, ale również grupy uprzywilejowane, których interesy reprezentowała. Pokrzywdzona zbiorowość identyfikowała się właśnie poprzez cechę różnicującą, co umożliwiało integrację oraz mobilizację polityczną i społeczną jednostek dyskryminowanych, a także ukierunkowanie poszczególnych działań na realizację wspólnych interesów.

Niestety władza, uznając postulaty wolnościowe, korygowała stan nierówności w zakresie niestanowiącym zagrożenia dla pozycji klasy rządzącej. Odpowiedź państwa uchwalającego ustawy miała z jednej strony ograniczać skalę wojen cywilnych i rewolucji, z drugiej zaś – zminimalizować straty ekonomiczne i polityczne związane z rekonstrukcją porządku społecznego wynikającą z uznania praw podstawowych marginalizowanych, a nawet wykorzystywanych grup. Działania te miały zatem na celu nie tylko ograniczenie nierówności, ale też utrzymanie władzy scentralizowanej przy grupach uprzywilejowanych przy jednoczesnym fragmentarycznym zwiększeniu dostępu do określonych praw lub możliwości grupom marginalizowanym.

W niniejszym podrozdziale starałam się ukazać, że aparat prawny może tworzyć, a następnie utrzymywać stan nierówności – w tym sensie prawo jest czasami bazą, a nie nadbudową istniejących procesów różnicowania. Proces walki o równość zakłada natomiast zlikwidowanie lub ograniczenie regulacji konstytuujących nierówności, co może prowadzić do reorganizacji życia społecznego bądź struktur majątkowych. Bodźcami, które wyzwalają kolejne etapy walki o równość, są: jawność kategoryzowania ze względu na zakwalifikowanie jednostki do danej grupy wyróżnionej ze społeczeństwa z uwagi na posiadaną cechę, wspólna tożsamość grupy

24:65944

pokrzywdzonych wynikająca z dostrzegalności cechy i niesprawiedliwość, która była odczuwalna zarówno przez członków grupy pokrzywdzonej, jak i resztę społeczeństwa.

Niezgoda na procesy różnicowania narasta w czasie i eskaluje w momencie identyfikowania się z grupą pokrzywdzoną różnicowaniem coraz liczniejszych jej przedstawicieli. Dostrzegalność procesu różnicowania, lecz także samej cechy uzasadniającej kategoryzację jest z tej perspektywy elementem kluczowym dla sprzeciwu społecznego. Zakaz dyskryminacji ma zatem znaczenie kontekstowe i ewolucyjne, gdyż jego funkcją jest uzupełnienie gwarancji równości w zakresie przyznanych praw i wolności. Tym samym w przypadku poszerzania katalogu tych praw lub reinterpretacji ich treści czy zakresu modyfikacjom winien ulegać również zakaz dyskryminacji.

2. Dyskryminacja w perspektywie społecznej

Zakaz dyskryminacji był reakcją na procesy polityczne, na które w ujęciu socjologicznym składało się nieusprawiedliwione okolicznościami, długotrwałe i celowe nierówne traktowanie ze względu na posiadanie określonej cechy[56]. Mobilizowały one do działania grupowego i sprzeciwu wobec wykluczenia oraz marginalizacji. Jak ukazałam w pierwszym podrozdziale, dyskryminacja jako pojęcie jest ściśle związana z pogwałceniem zasady sprawiedliwości, w związku z czym wykształcenie się zakazu dyskryminacji było nierozerwalnie związane z walką o równość względem prawa. Sprzeciw wobec niesprawiedliwości poprzedzało formowanie się tożsamości pokrzywdzonych grup. Podstawą nieuzasadnionego i niesprawiedliwego, a przy tym nierównego traktowania nie były bowiem indywidualne cechy osobowe bądź zachowania jednostek, ale identyfikacja grupowa, wiążąca się z odmową równouprawnienia całych kategorii lub grup społecznych[57].

Dyskryminacja jako zjawisko społeczne może występować w różnorodnych formach, a przy tym charakteryzować się złożoną i wieloaspektową strukturą. Wartą odnotowania analizę pojęciową dyskryminacji z perspektywy społecznej przeprowadził Martin Reisigl[58]. Na jej podstawie można przyjąć, że dyskryminacja zakłada istnienie: 1. sprawców należących do

[56] W. Burek, W. Klaus, *Definiowanie dyskryminacji w prawie polskim w świetle prawa Unii Europejskiej oraz prawa międzynarodowego*, „Problemy Współczesnego Prawa Międzynarodowego, Europejskiego i Porównawczego" 11, 2013, s. 75.
[57] Ch. Stangor, C.N. Macrae, M. Hewstone, *Stereotypy i uprzedzenia*, przeł. M. Majchrzak [*et al.*], Gdańsk 1999, s. 228.
[58] M. Reisigl, *Dyskryminacja w dyskursach*, „Tekst i Dyskurs – Text und Diskurs" 3, 2010, s. 27–61.

25:46887

konkretnej grupy społecznej lub kulturowej; 2. ofiar dyskryminacji; 3. zachowania lub procesu dyskryminacyjnego; 4. cechy, na podstawie której przebiega kategoryzacja (np. rasa, pochodzenie etniczne, płeć, orientacja seksualna); a także 5. porównywalnej grupy, w stosunku do której można stwierdzić niekorzystne różnicowanie[59]. W ramach uzupełnienia definicji autora można dodać jeszcze jedną cechę procesu dyskryminacji, mianowicie 6. trudność przerwania koła dyskryminacji, wynikającą z wykluczenia prawnego grup historycznie marginalizowanych. W niniejszym podrozdziale omówię wskazane przesłanki definicji dyskryminacji społecznej.

Jak wyjaśnia M. Reisigl[60], odwołując się do pierwszej przesłanki analizowanego pojęcia, dyskryminacja zakłada istnienie aktorów społecznych – sprawców. Samo występowanie różnic w społeczeństwie jest zjawiskiem typowym i nie prowadzi wprost do naruszenia praw czy wolności jednostki. Zasadniczo sprawcą dyskryminacji jest podmiot dokonujący krzywdzącego różnicowania. Może być ono jednak wynikiem szerszego problemu dominacji nad osobami odznaczającymi się pewną cechą (np. rasa, pochodzenie czy płeć)[61]. Tacy aktorzy będą dopuszczać się dyskryminacji nie tylko poprzez konkretne działanie, ale również przez utrzymywanie aparatu prawnego, który odgrywa konstytutywną rolę w umożliwianiu i podtrzymywaniu różnicowania. Przyjmuje się, że dyskryminujący z reguły dysponują władztwem różnicowania innych lub mogą sobie taką władzę przyznać. Struktura wykonywania władzy, cechująca się asymetrią między podmiotami dominującymi a podległymi władztwu, zazwyczaj uniemożliwia podporządkowanym przeprowadzenie procesów niekorzystnego różnicowania grupy dominującej z wyjątkiem sytuacji, kiedy przedstawiciele władzy są nieobecni i nie mogą jej sprawować[62].

Ten aspekt nierówności wpływający na charakter procesu dyskryminacji dostrzegał już John Stuart Mill[63]. Autor wyjaśniał, że prawa są stanowione przez władzę publiczną, działającą jako przedstawiciel określonych interesów, dzięki czemu podmioty reprezentowane w strukturach władzy mogą zgodnie ze swoim interesem umacniać lub tworzyć reguły postępowania albo praktyki pozwalające na zachowanie dominacji na skutek utrzymania stanu nierówności społecznej. Władza dążąca do utrzymania nierówności często jest przy tym budowana na

[59] *Ibidem*, s. 29.
[60] *Ibidem*.
[61] W. Klaus, *Prawo jako instrument mający zapewnić równość w społeczeństwie*, [w:] *Różni, ale równi. Prawo a równe traktowanie cudzoziemców w Polsce*, red. W. Klaus, Warszawa 2013, s. 11.
[62] C.F. Graumann, M. Wintermantel, *Discriminatory Speech Acts: A Functional Approach*, [w:] *Stereotyping and Prejudice: Changing Conceptions*, red. D. Bar-Tal *et al.*, New York 1989, s. 199.
[63] Por. C.R. Sunstein, *Sprzeciw w życiu społeczeństw*, przeł. J.S. Kugler, Warszawa 2006.

26:11361

podstawie teorii społecznych, antropologicznych czy politycznych racjonalizujących stosunki dominacji.

Uzasadnianie uprzedzeń prowadzących do dyskryminacji naturalnością istniejącego porządku będzie aktem przemocy symbolicznej, utrwalającym lub pogłębiającym niesprawiedliwość[64]. Władza różnicowania może się zatem przejawiać nie tylko w samym stosowaniu praktyk dyskryminujących, ale również w tworzeniu przestrzeni prawnej umacniającej takie procedery lub realnie im nieprzeciwdziałającej.

Drugim elementem konstytutywnym dla zrozumienia zjawiska dyskryminacji społecznej jest rozpoznanie grup dotkniętych niedozwolonym różnicowaniem. Jego ofiary często pochodzą z mniejszości społecznych lub grup marginalizowanych, a minorytyzacja oraz marginalizacja tych grup postępują wraz z procesem dyskryminacji[65]. Grup tych nie musi przy tym cechować liczebna mniejszość – dyskryminowani zwykle są wręcz porównywalną ilościowo grupą ze sprawcami (dzieje się tak m.in. w wypadku seksizmu, którego ofiarami zwykle są kobiety)[66].

Ustalenie kryteriów tworzenia marginalizowanych grup jest niezwykle istotne dla budowania definicji dyskryminacji, co dokładniej wyjaśnię w kolejnym podrozdziale. Wynika to z tego, że przyznanie grupom o określnych cechach ochrony przed dyskryminacją powoduje, iż grupy te zyskują pewną doniosłość na tle społecznym. Pokrzywdzeni z powodu zaliczenia ich do jednej grupy mogą się przy tym stać autorami zachowań dyskryminacyjnych wobec innej kategorii osób. Role społeczne przenikają się; co więcej, ludzie mogą reprezentować wiele cech równocześnie, a przez to być utożsamiani z wieloma grupami dyskryminowanymi. Różnicowanie może w związku z tym oddziaływać na jednostkę symultanicznie, ze względu na wiele aspektów, a co za tym idzie – ofiara będzie adresatem licznych form dyskryminacji.

Problem ten opisał w swojej analizie Reisigl, przywołując wykształcone w XX w. pojęcia dyskryminacji wielokrotnej (ang. *multiple discrimination*) oraz dyskryminacji krzyżowej (ang. *intersectional discrimination*), które służyły opisaniu szczególnych form dyskryminacji kobiet afroamerykańskich. Dyskryminacja tej grupy różniła się zarówno od dyskryminacji ze względu

[64] M. Reisigl, *op. cit*.
[65] T. Makkonen, *Hauptursachen, Formen und Folgen von Diskriminierung*, [w:] *Handbuch zur Rechtlichen Bekämpfung von Diskriminierung*, Helsinki 2003, s. 14.
[66] M. Reisigl, *op. cit*.

27:87163

na płeć (w wypadku dyskryminacji kobiet białych), jak i dyskryminacji ze względu na rasę (w przypadku dyskryminacji mężczyzn)[67].

Trzecim składnikiem pojęciowym dyskryminacji społecznej jest samo działanie dyskryminujące lub proces dyskryminujący. Należy jednak brać pod uwagę, że proces różnicowania sytuacji prawnej i faktycznej jednostek realizuje się równolegle na wielu płaszczyznach życia społecznego (np. edukacja, rekrutacja, dostęp do opieki zdrowotnej i innych usług publicznych) i może mieć źródło zarówno w jednostkowych decyzjach, jak i istniejącym ustawodawstwie, które tworzy stany nierówności lub sprzyja ich funkcjonowaniu. Efektem procesu dyskryminacji jest przy tym nie tylko różnicowanie jako takie, ale też naruszenie praw podstawowych bądź zasady sprawiedliwości polegające na ograniczeniu praw, wykluczeniu, widocznej segregacji, odmowie przyznania szans czy możliwości i równych praw. Omawiany proces jest zatem różnicowaniem z dodatkowym elementem pejoratywnym społecznie. Zasadnicze staje się w tym kontekście ustalenie granic, w których różnicowanie jest społecznie uzasadnione i użyteczne, i odróżnienie go od takiego różnicowania, które będzie kwalifikowane jako dyskryminacja. Istnieje wiele rozmaitych, często nawet wykluczających się, koncepcji związanych z ustalaniem, jakie różnicowanie nie narusza zasady równości. Wyróżnia się zarówno formalne i materialne ujęcia równości, jak i teorie równości szans oraz równości rezultatów, ściśle związane z koncepcją równości jako uznania i poszanowania różnorodności[68].

Równość może być przy tym rozumiana na dwa sposoby – jako równość w prawie i jako równość wobec prawa. Klasyfikacja ta koresponduje z tworzeniem i stosowaniem regulacji prawnych. Wojciech Sadurski jako podstawy uznania, kto powinien być równy lub taki jest, wskazywał słuszność oraz sprawiedliwość[69]. Nierówności społeczne mają bowiem tendencję do sukcesywnego pogłębiania się, odgrywając schemat błędnego koła. Wykluczenie prawne lub ograniczenie dostępu do dóbr w jednym obszarze pociąga więc za sobą nierówności w innych sferach życia (np. ograniczenia w edukacji wpływają na późniejszą marginalizację kobiet w procesach zatrudnienia). Z tej przyczyny kwestia równości wielokrotnie była już istotnym elementem walki politycznej, co przedstawiłam w poprzednim podrozdziale.

[67] S. Fredman, E.M. Szyszczak, *The Interaction of Race and Gender Discrimination*, [w:] *The Limits of Law*, red. B. Hepple, E.M. Szyszczak, London-New York 1992, s. 221.
[68] J. Maliszewska-Nienartowicz, *Karta Praw Podstawowych Unii Europejskiej. Komentarz*, red. A. Wróbel, Warszawa 2020, rozdz. 1.
[69] W. Sadurski, *Równość wobec prawa*, „Państwo i Prawo" 8–9, 1978, s. 59–60, 63–64.

28:47236

Wykluczenie grup upośledzonych, marginalizowanych lub prześladowanych oddziałuje bowiem na społeczeństwo jako całość, stopniowo prowadząc do jego rozpadu[70].

Czwartym elementem interesującej mnie tu definicji jest cecha wyróżniająca lub element, na którym można oprzeć dyskryminację. W ramach ustalenia tego, jakie różnicowanie nie narusza zasady równości, konieczne jest sprecyzowanie przedmiotu podobieństwa, a zatem wybór cechy. Podstawą takiej oceny mogą być pojęcia godności i równości każdego człowieka[71]. Cecha utożsamiana z przyczyną dyskryminacji może bowiem wynikać z negatywnych osądów, które wydajemy o innych, szczególnie obcych nam, grupach społecznych. W tym aspekcie istotną rolę odgrywają uprzedzenia i stereotypy, które mogą mieć korzenie w osobistych doświadczeniach bądź poglądach oraz osądach nabywanych w procesie socjalizacji. Z tą różnicą, że uprzedzenia, zwłaszcza te zasadzające się na ocenie podmiotowości (np. rasa), ról społecznych (np. płeć) czy też kwestii reprezentacji kulturowej (np. mniejszości etniczne), powstają irracjonalnie i w przeciwieństwie do stereotypów nie ustępują w przypadku większej ilości lub lepszej jakości informacji na temat, co do którego sprawca dyskryminacji był w błędnym przekonaniu[72]. Uprzedzenia wynikają bowiem z subiektywnego wyobrażenia rzeczywistości, ze względu na brak ich logicznego uzasadnienia uniemożliwiając dyskusję opartą na faktach.

Stygmatyzacja będąca następstwem utrwalonych uprzedzeń panujących w relacjach społecznych wymyka się regulacjom prawnym i z tej przyczyny nie może być efektywnie ograniczana jedynie na drodze legislacyjnej[73]. Utrwalone stygmaty bywają wręcz interpretowane jako swego rodzaju odchylenie od pozytywnej „normalności”[74]. Uprzedzenia są trudne do przełamania, a proces ich mitygowania jest skomplikowany i długotrwały. Brak reakcji władzy czy społeczne przyzwolenie na emanację w sferze publicznej uprzedzeń ugruntowuje je, nadając różnicowaniu narrację słuszności i przeradzając procesy negatywnie nacechowanej stygmatyzacji w dyskryminację. To zaś zwrotnie pogłębia cykl wykluczenia marginalizowanych grup.

[70] J. Maliszewska-Nienartowicz, *op. cit.*, rozdz. 3, nb. 8.
[71] W. Burek, W. Klaus, *op. cit.*, s. 77.
[72] E.F. Borgatta, R.J.V. Montgomery, *Encyclopedia of Sociology*, New York 2000, s. 688–889, 693–694.
[73] A. Winiarska, W. Klaus, *Dyskryminacja i nierówne traktowanie jako zjawisko społeczno-kulturowe*, „Studia BAS” 2, 2011, nr 26, s. 18.
[74] M.L. DeFleur, *“Stigma: Notes on the Management of Spoiled Identity” by Erving Goffman. Englewood Cliffs, New Jersey: Prentice-Hall, 1963. 147 pp. Cloth, $4.50; paper, $1.95*, „Social Forces” 43, 1964, nr 1, s. 127–128.

29:38622

Nie każde różnicowanie, nawet oparte na bezpodstawnych uprzedzeniach, będzie jednak wyczerpywało znamiona dyskryminacji (np. uprzedzenia związane z kolorem włosów nie muszą wszak prowadzić do dyskryminacji). W związku z tym cecha wyróżniająca daną grupę i będąca przyczyną niekorzystnego różnicowania powinna mieć istotny wymiar społeczny (logikę tworzenia katalogu cech chronionych szczegółowo omawiam w kolejnym podrozdziale).

Piątym elementem konstruktywnym dyskryminacji jest odniesienie sytuacji osób z daną cechą do grupy jej pozbawionej. Oznacza to przeprowadzenie porównania sytuacji prawnej i faktycznej grup, które z jednej strony łączy podobieństwo, z drugiej zaś odróżnia posiadanie określonej cechy, oraz ustalenie granicy uzasadnionego różnicowania. Taka analiza jest zasadniczo prosta, gdy dyskryminacja bezpośrednia opiera się na jednej cesze. Jednakże kręgi podmiotów dysponujących cechami chronionymi przenikają się (np. ciemnoskóre kobiety, kobiety z niepełnosprawnościami), przez co analiza porównawcza może się okazać niezwykle trudna lub nawet niemożliwa. Problem ten dotyczy przede wszystkim przypadków dyskryminacji pośredniej lub wielokrotnej, jak również emergentnej. Jednostka dyskryminowana nie jest bowiem oczywistym przykładem egzemplarycznym większej jednolitej społeczności, co prowadzi do niemożliwości ustalenia i zidentyfikowania porównywalnej grupy. Dodatkowo wielość cech utrudnia przesądzenie, na podstawie której z nich jednostka została pozbawiona praw lub w nich ograniczona. Brak odniesienia uniemożliwia wykazanie gorszego położenia na skutek przeprowadzonego procesu różnicowania.

Uzupełniając pięć punktów opisanych przez Martina Reisigla, należy za Witoldem Klausem[75] wskazać, że obserwowany jest również szósty element procesu dyskryminacji, którym jest wykluczenie prawne. W socjologii zjawisko to jest rozumiane jako wtórna wiktymizacja społeczna – jest bowiem bezpośrednio związane z położeniem finansowym i kulturowym pokrzywdzonej jednostki[76]. Dysproporcjom ekonomicznym często wszak towarzyszą deficyty polityczne, wynikające z niedostatecznej reprezentacji i przejawiające się w tym, że władza nie realizuje postulatów równościowych w kontekście ochrony praw wszystkich grup społecznych. Zostało to już ukazane w toku analizy historycznej walki o równość.

[75] W. Klaus, *op. cit.*, s. 14.
[76] A. Turska, *Prawo i wykluczenie. Studium empiryczne*, Warszawa 2018.

30:49200

Co więcej, nie wszyscy członkowie grup wykluczonych są w stanie w wystarczającym stopniu rozpoznawać i korzystać ze swoich demokratycznych praw. Prawo tworzone jest bowiem przez elity polityczne, reprezentujące określone interesy grup dominujących nad wykluczonymi, przez co zarówno władza, jak i jej elektorat w celu zachowania istniejącego porządku mogą utrwalać stereotypy dotyczące członków grup marginalizowanych[77].

Wykluczenie prawne, które napędza procesy dyskryminacji, nie jest przy tym jedynie efektem procesu tworzenia i stosowania prawa, ale również uwidocznia deficyt kompetencji, pozycji społecznej i możliwości finansowych w zakresie przeciwstawiania się krzywdzącym decyzjom. Zjawisko to przybiera postać pogłębiającego się cyklu. Prawo pomija pewne grupy, ponieważ władza na etapie procesów legislacyjnych nie słyszy głosu ich reprezentantów bądź nie dostrzega ich potrzeb, a nie może lub nie chce ich usłyszeć, gdyż wykluczeni nie mają rzeczywistej siły lub możliwości dojścia do głosu w celu zmiany prawa[78]. Społeczeństwo, podzielone na grupy mające możliwości i reprezentację kosztem grup wykluczonych, niedoreprezentowanych lub ograniczonych w prawach, koroduje[79]. Dostrzegalne i utrwalone różnicowanie może bowiem rodzić napięcia społeczne wynikające z poczucia niesprawiedliwości, które prowadzą do rewolucyjnych przewrotów. Nieprzeciwdziałanie utrwalaniu się uprzedzeń prowadzących do dyskryminacji jest zatem źródłem reprodukcji niesprawiedliwości i zagrożenia demokracji. Ostatecznie nie można wykluczyć łączenia się w grupy osób podobnie odrzuconych i pomijanych przez państwo, a przez to funkcjonujących na granicy społeczeństwa, które to grupy w ramach tworzonych społeczności zaczną budować swoją tożsamość adekwatnie do wyobrażeń wynikających ze stereotypów[80].

Poczuciu odrzucenia i wykluczenia z systemu towarzyszy wzrost poczucia odrębności przedstawicieli poszczególnych grup, a w konsekwencji postrzeganie innych jako obcych. Ta

[77] Ph. Alston, *Report of the Special Rapporteur on Extreme Poverty and Human Rights*, United Nations General Assembly 2015.
[78] W. Klaus, *op. cit.*, s. 15.
[79] *Ibidem*.
[80] „Uprzedzenia i dyskryminacja prowadzą do powstawania dzielnic wykluczonych. Przegląd amerykańskich badań w tym zakresie, obejmujących przede wszystkim zagadnienia związane z rasą, prowadzi do wniosków, że proces ten dokonuje się już na etapie wynajmowania mieszkań. Okazuje się bowiem, że dla przedstawicieli pewnych grup tylko niektóre lokale czy lokalizacje są dostępne. I znajdują się one w tzw. gorszych dzielnicach. Tak powstają getta. Warto przyjrzeć się skutkom takiej przestrzennej dyskryminacji. Wyniki badań pokazują, że osoby mieszkające w enklawach skupiających wykluczonych mają mniejsze szanse na poprawę warunków swojego życia, a nawet na życie zgodne z oczekiwaniami pozostałej części społeczeństwa. Trudniej im znaleźć pracę, bowiem w dzielnicy, którą zamieszkują, najczęściej jej nie ma, więc muszą dojeżdżać do innego regionu miasta. Dzielnice te są gorzej skomunikowane z resztą miasta, więc czas dojazdu do pracy jest znacznie dłuższy (nawet jeśli formalnie droga do pracy nie jest dłuższa niż dla przeciętnego członka społeczeństwa)" – A. Winiarska, W. Klaus, *op. cit.*, s. 18.

31:11127

„obcość" staje się przedmiotem niezrozumienia, a więc i lęku. Za Z. Baumanem można przyjąć, że nawet jeżeli obcy nie stają się obiektem jawnej agresji i nie są otwarcie obrażani, ich obecność wprawia w zakłopotanie[81]. Jednocześnie może ona intensyfikować powstawanie uprzedzeń, wpływających na dalsze relacje między grupami. Ład społeczny wykształcony na stygmatyzacji powoduje, że niekorzystne różnicowanie i powodowany nim stan niesprawiedliwości jest racjonalizowany i nie budzi refleksji u jego aktorów.

Ewolucyjna forma podziałów społecznych, wynikająca z nieodłącznej kategoryzacji ludzi (również na postawie irracjonalnych uprzedzeń i przypisywanych im stygmatów), potęguje ryzyko zaistnienia procesów dyskryminacji w nowożytnych demokracjach i przyczynia się do tworzenia nowych krajobrazów nierówności. W tym aspekcie dyskryminację można uznać jednocześnie za przyczynę i konsekwencję wykluczenia i marginalizacji. Walka o równość nie zakończyła się bowiem wraz z przyjęciem aktów prawnych, a aktualne procesy dyskryminacji mogą ponownie doprowadzić do takiego poziomu niesprawiedliwości, który w efekcie przyniesie sprzeciw o charakterze rewolucyjnym. Analiza metod ograniczenia wykluczenia i marginalizacji ma zatem wymiar aktualny i istotny z perspektywy zapewnienia efektywnej ochrony praw człowieka.

3. Pojęcie dyskryminacji

Powszechne rozumienie dyskryminacji zostało ukształtowane przede wszystkim przez sposób sformułowania zakazu różnicowania istniejący w prawie antydyskryminacyjnym[82]. Akty prawne, odnosząc się do pojęcia dyskryminacji, na ogół opisują katalogi cech chronionych, na podstawie których kategoryzowanie w dostępie do praw i możliwości jest zakazane. Takie budowanie definicji rodzi jednak wiele pytań, albowiem różnicowanie ludzi samo w sobie nie jest moralnie ani prawnie kontrowersyjne. Dlaczego zatem pewne grupy zasługują na ochronę, a inne nie? Jakimi kryteriami powinien się kierować legislator, zakazując różnicowania, opierając się na takich cechach jak płeć, rasa, wiek czy religia zamiast wzrostu lub koloru włosów bądź upodobania do gry w piłkę? A jeżeli takie cechy również mogą prowadzić do dyskryminacji, to czy państwo powinno podchodzić do każdej z tych form dyskryminacji w

[81] Z. Bauman, *Płynne czasy. Życie w epoce niepewności*, Warszawa 2007, s. 119.
[82] S. Moreau, *Discrimination Law and the Freedom to Live a Good Life*, „Law and Philosophy" 35, 2016, s. 517.

32:28027

podobny sposób? Dlaczego nie istnieje jeden uniwersalny katalog cech chronionych stworzony do celów każdego typu różnicowania?

Poszukiwanie odpowiedzi na wskazane pytania jest istotne badawczo, ponieważ pozwala ustalić, czy i w jakim zakresie prawo antydyskryminacyjne opierające się na pojęciu dyskryminacji podąża za rzeczywistymi relacjami dominacji mającymi charakter ewolucyjny. Założenie prezentowane przez część teoretyków[83], że prawo antydyskryminacyjne jest pozbawione wspólnej podstawy normatywnej uzasadniającej przypisanie tylko niektórym zachowaniom różnicującym charakteru dyskryminacji, znacząco zubożałoby dyskusję i zamykało ją na efektywny rozwój legislacji przeciwdziałającej marginalizacji społecznej i wykluczeniu. W związku z tym zasadne zdaje się przeprowadzenie teoretycznej analizy prawa antydyskryminacyjnego i pojęć wykorzystywanych na tym gruncie w poszukiwaniu wiedzy o przedmiocie regulacji.

Na początku trzeba podkreślić, że nie ma jednej powszechnie akceptowanej prawniczej definicji dyskryminacji[84]. Pojęcie to obejmuje bowiem szeroką gamę rozbieżnych przypadków różnicowania, wykraczających poza klasyczne granice etyki stosowanej i teorii politycznej: od sprawiedliwości rozdzielczej, sądownictwa i etyki medycznej po feminizm, wielokulturowość, prawa liberalne czy tożsamość płciową[85]. W filozofii krajowej refleksja nad terminem „dyskryminacja" nie jest pogłębiona, a pojęcie to wykorzystywane jest nieprecyzyjnie. Może to powodować nieporozumienia oraz prowadzić do błędnych wniosków. W niniejszym podrozdziale przeprowadzę w związku z tym analizę autorską samego pojęcia dyskryminacji z uwzględnieniem dyskusji co do tego, czy do istnienia dyskryminacji konieczne jest operowanie na cechach chronionych, a jeżeli tak, jaka jest logika leżąca u podstaw tworzenia katalogu takich cech. Rozważę również, czy dyskryminacja może istnieć w oderwaniu od cech chronionych.

Zakaz dyskryminacji jest elementem prawa równościowego, co oznacza, że prawo antydyskryminacyjne jest swoistym metaprawem, które ma zabezpieczać równe korzystanie z praw i wolności – omawiany zakaz miał zabezpieczać równouprawnienie jednostek. Z tego powodu dyskryminacja ma historycznie akcesoryjny charakter, tj. może zaistnieć w związku z

[83] Por. P.S. Shin, *Is There a Unitary Concept of Discrimination?*, [w:] *Philosophical Foundations of Discrimination Law*, red. D. Hellman, S. Moreau, Oxford 2013, s. 163–181.
[84] W. Vandenhole, *Non-Discrimination and Equality in the View of the UN Human Rights Treaty Bodies*, Oxford 2005, s. 33.
[85] F.K. Thomsen, T. Khaitan, F. Schauer, *The Routledge Handbook of the Ethics of Discrimination*, red. K. Lippert-Rasmussen, New York-London 2017, s. 21.

33:65588

nierównym dostępem do prawa lub wolności, czy to na poziomie stosunków prywatnych, czy na poziomie systemowym[86]. Samo w sobie różnicowanie jest nieodrodnym elementem każdego procesu decyzyjnego, w związku z czym bez odpowiedniego kontekstu nie powoduje niesprawiedliwej dystrybucji możliwości czy zasobów. Dlatego też nie każde różnicowanie jest zakazane, mimo że większość procesów decyzyjnych prowadzi do wyborów ograniczających prawa lub wolności jednostki. Dyskryminacja wymaga zatem dodatkowego elementu oprócz samego procesu różnicowania[87].

Wobec konieczności rozgraniczenia dozwolonego różnicowania od penalizowanej dyskryminacji rozpoczęła się w filozofii ożywiona debata. Frej Klem Thomsen[88] analizował, czy dyskryminacja powinna być specyficzna dla danego celu różnicowania, tzn. czy w definicji dyskryminacji powinno znajdować się zastrzeżenie, że jest ona skierowana przeciwko konkretnym grupom, a jeśli tak, to w jaki sposób powinniśmy określić grupy, które mogą być obiektem dyskryminacji. W tym zakresie można się odwołać do prezentowanej w niniejszej pracy historycznej walki o równość. W jej toku wykształciły się bowiem cechy chronione prawem antydyskryminacyjnym, takie jak rasa, płeć, religia czy orientacja seksualna. Cechy te odnoszą się przy tym wprost do ewolucyjnego poszerzania się praw podstawowych i obejmowania nimi wykluczonych uprzednio grup społecznych. Jak zauważa Tarunabh Khaitan[89], aby zyskać miano chronionej, dana cecha musi mieć charakter osobisty, tj. dotyczyć lub pomagać w opisie ludzi – w przeciwieństwie do rzeczy, emocji czy też stanów umysłu. Jest to zgodne z wskazanymi wcześniej koncepcjami prawnonaturalnymi, pojmującymi równość jako element godności człowieka.

Z kolei Kasper Lippert-Rasmussen[90] analizuje konieczność przypisania jednostce odpowiedzialności za posiadaną cechę oraz istnienie nierozerwalnej więzi między jednostką a cechą[91]. Cechy osobowe służące opisywaniu ludzi wyróżniamy, kierując się porządkiem uniwersalnym[92], co umożliwia nam wyszczególnienie społeczności identyfikujących się

[86] W ostatnich latach coraz więcej uwagi poświęca się innemu obszarowi dyskryminacji, mianowicie temu, w którym dyskryminacja odnosi się do statusu ludzi jako wiedzących; por. *ibidem*, s. 3.
[87] *Ibidem*, s. 2.
[88] *Ibidem*, s. 25.
[89] T. Khaitan, *Prelude to a Theory of Discrimination Law*, [w:] *Philosophical Foundations*…, s. 138–162.
[90] K. Lippert-Rasmussen, *The Badness of Discrimination*, „Ethical Theory and Moral Practice" 9, 2006, nr 2, s. 169.
[91] Podczas gdy rasa, kolor skóry, pochodzenie narodowe lub społeczne oraz okoliczności urodzenia są niezmienne, język, religia, poglądy polityczne lub inne rodzaje poglądów, własności i innego statusu takie nie są. Co więcej, istnieją inne cechy, które są niezmienne (np. rozmiar buta), które zazwyczaj nie znajdują się na liście istotnych cech; por. B. Heinrichs, *Discrimination in the Age of Artificial Intelligence*, „AI & Society" 37, 2021, s. 145.
[92] M.P. Foran, *Grounding Unlawful Discrimination*, „Legal Theory" 28, 2022, s. 10.

34:42420

poprzez te cechy. Konsekwencją ustalenia grup byłaby zatem możliwość rozstrzygnięcia o sposobie formowania cech. Konieczne staje się w związku z tym poznanie logiki formowania grup objętych ochroną przeciwdyskryminacyjną, a w konsekwencji odpowiedź na pytanie, dlaczego określone cechy mogą być podstawą dyskryminacji, a inne nie.

Thomsen wskazał, że istotność grup opiera się na łączącej je tożsamości, polegającej na politycznym niedoreprezentowaniu, społecznym wykluczeniu i złej sytuacji materialnej, co historycznie doprowadziło do marginalizacji, podporządkowania i zniewolenia danej grupy społecznej[93]. Lippert-Rasmussen, prowadząc rozważania na temat historycznie wykształconych cech chronionych prawem antydyskryminacyjnym (takich jak płeć, rasa, pochodzenie), zasugerował natomiast podejście kontekstowe, posadowione na analizie wagi grupy pod kątem jej społecznej doniosłości i wpływu na strukturę interakcji społecznych[94]. W tym zakresie należy zbadać, na ile różnicowanie tak uformowanej grupy będzie pogłębiało marginalizację i wykluczenie historycznie uciskanych społeczności oraz w jakim kontekście to różnicowanie ma miejsce[95]. Co do zasady bowiem różnicowanie ze względu na kolor skóry czy płeć będzie potęgowało zło, jakie wyrządziły uprzedzenia istniejące w społeczeństwie na tym tle od czasów niewolnictwa czy momentu uprawomocnienia patriarchatu. Jednocześnie różnicowanie to może być dopuszczalne, o ile okoliczności procesu różnicowania będą wskazywały na inne podłoże stawianych kryteriów (np. casting do roli kobiecej lub postaci historycznej).

Bert Heinrichs, analizując wskazane założenia, uznaje za Thomsenem, że w zakresie, w jakim chcemy potępić praktyki dyskryminacyjne, nie ma oczywistego powodu, dla którego dyskryminacja idiosynkratyczna nie byłaby zła z tych samych powodów, dla których dyskryminacja paradygmatyczna jest moralnie odrażająca (przynajmniej czasami), nawet jeśli będzie mniej powszechna, a przez to zło przez nią wyrządzane będzie mniejsze[96]. Teoria kontekstowa, chociaż krytykowana w literaturze[97], wydaje się jednak lepiej dostrzegać specyfikę różnorodności i wieloaspektowości procesów decyzyjnych, a także trafniej odpowiadać na możliwe przemiany w zakresie znaczenia konkretnych grup dla ogółu społeczności. Zasadniczo dane cechy powinny prowadzić do jednoznacznego określenia

[93] F.K. Thomsen, T. Khaitan, F. Schauer, *op. cit.*, s. 25.
[94] K. Lippert-Rasmussen, *op. cit.*, s. 169.
[95] *Ibidem*.
[96] B. Heinrichs, *op. cit.*, s. 145.
[97] Krytycznie: F.K. Thomsen, *But Some Groups Are More Equal than Others: A Critical Review of the Group-Criterion in the Concept of Discrimination*, „Social Theory and Practice" 39, 2013, nr 1, s. 120–146.

35:60640

reprezentantów grupy. Tak ustalony katalog cech uniemożliwia przyjęcie dyskryminacji z uwagi na sposób zachowania czy dane behawioralne. Rozważenia wymaga jednak problem nakładających się lub krzyżujących się cech, kiedy grupy przenikają się lub nakładają na siebie, tworząc nowe rodzaje dyskryminacji.

Do różnicowania na podstawie cech konieczna jest rzeczywista władza społeczna, którą mają zarówno instytucje, jak i organy publiczne. Tak rozumianą dyskryminację rozpoznajemy w niesprawiedliwym procesie lub w efekcie różnicowania pomiędzy podobnymi grupami ludzi, ponieważ niekorzystne traktowanie społeczności dyskryminowanej ma miejsce w przestrzeni publicznej i jest dostrzegalne również przez jednostki niedoświadczające dyskryminacji.

Dlatego też należy za Thomsenem[98] postawić kolejne pytanie – mianowicie, czy dyskryminacja jest pojęciem opisowym, czy moralnym? Można za częścią teoretyków[99] przyjąć, że dyskryminacja kwalifikowana jest negatywną oceną moralną różnicowania, taką jak niesprawiedliwość czy etyczna niedopuszczalność zachowania. Takie podejście ma swoje źródła w koncepcji równości Immanuela Kanta[100], zgodnie z którą dyskryminacja jest niemoralna, gdy różnicowanie opiera się na cechach moralnie nieistotnych, a za takie historycznie uznawano rasę czy płeć. Deborah Hellman[101] zauważa przy tym, że już samo nazywanie czegoś dyskryminacją oznacza krytyczną ocenę procesu różnicowania[102]. Pogląd ten jest słuszny w kontekście bezprawnego charakteru dyskryminacji uregulowanej w przepisach prawa. Heinrichs przybliża ten spór, wskazując, że część teoretyków[103] nie utożsamia dyskryminacji z moralnością, gdyż twierdzi, iż nie można z dyskryminacji wnioskować, że coś jest niesprawiedliwe lub moralnie niedopuszczalne[104].

Podsumowując zaprezentowaną w tym podrozdziale analizę, dochodzę do wniosku, że dyskryminacja jako odpowiedź na rzeczywisty problem dominacji nie musi być opisywana katalogiem cech chronionych, przy czym takie rozumienie zakazu różnicowania zostało ugruntowane w domenie publicznej na tle walk o równość. Katalog aktualnych cech chronionych był tworzony na podstawie przeglądu grup historycznie marginalizowanych i ich drogi do równouprawnienia. Definicja dyskryminacji przesądza o tym, że jest ona

[98] F.K. Thomsen, T. Khaitan, F. Schauer, *op. cit.*, s. 25.
[99] F.K. Thomsen, *op. cit.*
[100] Kant twierdził, że wszyscy ludzie są równi, zasługują na jednakowy szacunek i powinni być oceniani jedynie na podstawie swoich moralnych wyborów.
[101] D. Hellman, *When Is Discrimination Wrong?*, Cambridge, MA-London 2011, s. 2.
[102] Por. B. Heinrichs, *op. cit.*, s. 145.
[103] K. Lippert-Rasmussen, *op. cit.*
[104] B. Heinrichs, *op. cit.*, s. 145.

36:42259

metaprawem, czyli jest połączona z innym prawem, którego jednostka jest pozbawiona lub w którym jest ograniczona. Materialna treść tego prawa, którego na skutek dyskryminacji jednostka jest pozbawiana, istnieje odrębnie od zasady równości. Sugeruję przy tym, aby za prawniczym rozumieniem dyskryminacji uznawać ją za mającą moralnie negatywną kwalifikację, chociażby z racji naruszenia przez sprawcę powszechnie obowiązujących przepisów.

W dobie rozwoju technologii dyskryminacja straciła jednak dla aktorów procesów dyskryminacyjnych walor widzialności. Co więcej, zakaz dyskryminacji staje się nieoczywisty, gdy natrafiamy na nowe, domniemane przypadki dyskryminacji i musimy ustalić, czy rozszerzają one listę form penalizowanego różnicowania. Analizując aktualne procesy różnicowania, należy również wskazać na łączenie i zwielokrotnianie podstaw kategoryzowania ludzi oraz tworzenie nowych grup identyfikujących swoich członków na podstawie cech społecznie nieistotnych. Dyskryminacja nie musi się bowiem wiązać z intencją wykluczenia, a mimo to może prowadzić do społecznie negatywnych skutków.

4. Formy dyskryminacji

Rozważania na temat treści zakazu dyskryminacji doprowadziły do licznych pojęciowych rozróżnień w zakresie możliwych form tego zjawiska. Na tym tle uformowały się definicje dyskryminacji bezpośredniej oraz pośredniej, jak również organizacyjnej i instytucjonalnej. Podobnie jak w wypadku ogólnej definicji dyskryminacji także w odniesieniu do jej form pojawiają się spory. Formy dyskryminacji w przeciwieństwie do cech chronionych nie są bowiem elementem definicji samego zjawiska, jednak – tak jak w przypadku cech chronionych – ich zrozumienie pomaga wyznaczyć granicę dozwolonego różnicowania.

W pierwszej kolejności wyjaśnienia wymaga pojęcie dyskryminacji bezpośredniej. Ma ona miejsce wówczas, gdy różnicowanie jest oparte wprost i jedynie na chronionej cesze lub gdy różnica w traktowaniu jest bezpośrednio posadowiona na takim kryterium[105]. Oznacza to, że sprawca nie traktowałby innych grup w sposób niekorzystny, gdyby nie reprezentowały one konkretnej cechy chronionej. Dyskryminacja bezpośrednia, wykształcona na gruncie prawa europejskiego w przeciwieństwie do systemów *common law*, nie wymaga od agenta

[105] J. Maliszewska-Nienartowicz, *op. cit.*, preprint 21, nb. 31.

37:10657

umyślności w działaniu[106]. Okoliczność tę należy ocenić pozytywnie, albowiem wielokrotnie to nieuświadomione uprzedzenia czy nienacelowane na dyskryminację założenia gospodarcze mogą odpowiadać za procesy dyskryminacyjne. Uzupełniając zaprezentowane do tej pory rozważania, można za Lippertem-Rasmussenem dodać, że za dyskryminację bezpośrednią należy uznać też społeczną obojętność na pokrzywdzenie interesów grup niewystarczająco reprezentowanych w strukturach władzy[107].

Następnie trzeba wyjaśnić znaczenie dyskryminacji pośredniej. Pojęcie to wykształciło się w toku stosowania prawa antydyskryminacyjnego jako odpowiedź na deficyty treści i formy dyskryminacji bezpośredniej. Europejski Trybunał Praw Człowieka uznał bowiem, że mimo niewypełnienia przesłanek dyskryminacji bezpośredniej skutek procesu różnicowania może mieć analogiczny charakter. W tych okolicznościach Trybunał wywiódł, że w przypadku gdy ogólna polityka lub środek ma nieproporcjonalnie szkodliwe skutki dla określonej grupy, nie jest wykluczone, że może to zostać uznane za dyskryminację, mimo że nie jest konkretnie skierowane na tę grupę[108]. Dyskryminacja pośrednia zakazuje zatem stosowania z pozoru neutralnych przepisów, kryteriów czy praktyk, jeżeli wpływają one w szczególnie niekorzystny sposób na przedstawicieli grupy reprezentującej cechę chronioną pośrednio. Jednocześnie dyskryminacja pośrednia, podobnie jak bezpośrednia, może zaistnieć niezależnie od intencji sprawcy.

Koncepcja tego typu dyskryminacji jest wyjątkowo doniosła w tych systemach, które przewidują penalizację różnicowania jedynie na podstawie zamkniętego katalogu cech chronionych[109]. Oznacza to, że analizie poddawany jest wynik procesu różnicowania, a nie jego przebieg. Jeżeli różnicowanie następuje z uwagi na cechę inną niż chroniona, ale krzywdzący skutek procesu realnie dotyka grupę reprezentującą cechę chronioną, to owo działanie będzie naruszało zakaz dyskryminacji.

Koncepcja dyskryminacji pośredniej może się również okazać niezwykle istotna z perspektywy rozwoju automatycznych systemów decyzyjnych, których nieprzejrzystość epistemiczna utrudnia zrozumienie i nadzór nad procesami różnicowania, o czym szerzej będzie mowa w dalszej części pracy.

---

[106] Por. wyrok Europejskiego Trybunału Praw Człowieka (dalej: ETPCz) z 28.05.1985 r. w sprawie *Abdulaziz, Cabales i Balkandali v. Wielka Brytania*, skarga nr 9815/82.
[107] K. Lippert-Rasmussen, *Born Free and Equal? A Philosophical Inquiry into the Nature of Discrimination*, Oxford 2013.
[108] Wyrok ETPCz z 4.05.2001 r. w sprawie *Shanaghan v. Wielka Brytania*, skarga nr 37715/97.
[109] J. Maliszewska-Nienartowicz, *op. cit.*, preprint 21, nb. 41.

38:90429

Na tle definicji dyskryminacji pośredniej pojawił się jednak spór podający w wątpliwość, czy rzeczywiście w przypadkach tak rozumianej dyskryminacji pośredniej mamy do czynienia ze spełnieniem konstytutywnych przesłanek dyskryminacji. Krytycy koncepcji dyskryminacji pośredniej, m.in. Iris M. Young[110] i Matt Cavanagh[111], wskazują, że nierówna dystrybucja dóbr bądź ograniczenie w dostępie do nich może wynikać także z innych przyczyn, które nie będą mogły zostać zakwalifikowane jako dyskryminujące. Dostrzegają oni zatem problem wykluczenia czy też marginalizacji i potwierdzają jego niekorzystny wpływ na społeczeństwo, jednocześnie zaprzeczając, że takie wykluczenie będzie zakazane na podstawie przepisów antydyskryminacyjnych.

Jeżeli akty dyskryminacji pochodzą od organizacji, naruszenia zasady równości przypisuje się podmiotom zbiorowym i określa je mianem dyskryminacji organizacyjnej. Tego typu dyskryminacja może mieć charakter bezpośredni lub pośredni. Należy ją jednak odróżnić od dyskryminacji strukturalnej, nazywanej również instytucjonalną. Dyskryminacja strukturalna dotyczy bowiem zasad i przepisów konstytuujących oraz regulujących główne obszary życia społecznego, takie jak własność czy prawa polityczne. Głównym elementem takiego procesu nie jest intencja dyskryminacji, ale bierność zmierzająca do utrzymania efektu nierówności poprzez dopuszczenie do dominacji nad grupami historycznie marginalizowanymi[112]. Taka tendencja władzy publicznej do utrzymywania porządku opartego na nierównościach społecznych i stosunkach dominacji uzasadnia walkę o równość na przestrzeni wieków. Historyczne uwarunkowanie zakorzenione w ludzkiej skłonności do kategoryzowania ludzi na zasadach zaliczania ich do grup bliższych (znanych) i dalszych (obcych) nadaje dyskryminacji systemowej ciągle aktualny wymiar, albowiem do dziś grupy dominujące tworzą i narzucają mniejszościom postrzeganym jako obcy własne wzorce społeczne.

Kolejną formą dyskryminacji jest dyskryminacja wielokrotna. Powstaje ona na gruncie różnych cech chronionych, które istnieją i oddziałują na sytuację człowieka w rozmaitych momentach i niezależnie od siebie[113]. Przykładowo kobieta z niepełnosprawnością może być dyskryminowana z powodu swojej płci w procesie rekrutacji, a ze względu na swoją

[110] I.M. Young, *Justice and the Politics of Difference*, Princeton, NJ 2022.
[111] M. Cavanagh, *Against Equality of Opportunity*, Oxford 2002.
[112] F.L. Pincus, *Discrimination Comes in Many Forms: Individual, Institutional, and Structural*, „The American Behavioral Scientist" 40, 1996, nr 2, s. 190.
[113] T. Makkonen, *Multiple, Compound and Intersectional Discrimination: Bringing the Experiences of the Most Marginalized to the Fore*, Åbo Akademi University, Institute for Human Rights, Turku 2002, s. 10, https://www.abo.fi/wp-content/uploads/2018/03/2002-Makkonen-Multiple-compound-and-intersectional-discrimination.pdf (dostęp: 4.12.2021).

39:32253

niepełnosprawność w kontekście dostępu do instytucji publicznych, gdy budynek urzędu nie jest przystosowany do inwalidów[114].

Podobną formą dyskryminacji jest dyskryminacja krzyżowa; ona również jest złożona, jako że proces różnicowania odbywa się tu na gruncie wielu cech chronionych, które istnieją i oddziałują na sytuację dyskryminowanego równocześnie i równolegle, na jednym i tym samym polu i w tożsamej sytuacji społecznej[115]. Przykładem takiej dyskryminacji jest nieuzasadnione nakazanie sterylizacji kobiet z niepełnosprawnościami. Tego typu dyskryminacji nie doznają kobiety w ogólności ani niepełnosprawni mężczyźni. Należy przy tym podkreślić, że przypadki dyskryminacji złożonej nabierają znaczenia w kontekście poszukiwania grup porównywalnych, możliwości identyfikacji ofiary dyskryminacji z konkretną społecznością, a także ustalania wagi poszczególnych cech chronionych dla dyskryminującego efektu różnicowania. Mimo oczywistych walorów omówionych koncepcji ich stosowanie w praktyce napotyka bowiem na wiele trudności.

Warto na końcu wspomnieć o szczególnym podtypie dyskryminacji, jakim jest dyskryminacja statystyczna. Polega ona na dokonywaniu różnicowania w celu racjonalnego, efektywnego zarządzania przy świadomości deficytu informacji. Dyskryminacja statystyczna jest zatem narzędziem zarządzania ryzykiem, które może zostać uznane za racjonalny i społecznie neutralny sposób podejmowania decyzji.

Do stosunkowo niedawna rzadko była ona przy tym tematem dyskusji akademickich, jako że nie doszukiwano się w niej zasadniczego zagrożenia dla stosunków społecznych[116]. Frederick Schauer[117] jako przykład dyskryminacji statystycznej przywołuje jednak dyskryminację ze względu na wiek. Jako że wiek jest cechą chronioną w kontekście równouprawnienia w toku rekrutacji, a przy tym wiarygodnym wskaźnikiem problemów zdrowotnych, nie można odmówić celowości wyznaczaniu maksymalnego wieku dla kandydatów na pilotów komercyjnych linii lotniczych. Jednocześnie powszechnie akceptowany jest zakaz sprzedaży alkoholu osobom niepełnoletnim. W obu przypadkach różnicowanie przebiega na osi kryteriów chronionych, przy czym mimo ograniczenia w dostępie do możliwości różnicowanie jest oczywiście uzasadnione i społecznie pożądane. Autor rozważa natomiast bardziej dyskusyjne przykłady dotyczące uprzedzeń na gruncie równouprawnienia płci[118], albowiem do

[114] *Ibidem*.
[115] *Ibidem*, s. 11.
[116] F.K. Thomsen, T. Khaitan, F. Schauer, *op. cit.*, s. 51.
[117] *Ibidem*, s. 44.
[118] *Ibidem*.

40:76158

dyskryminacji statystycznej może również dojść, gdy na skutek statystycznie wyższego prawdopodobieństwa wyklucza się kobiety z procesu rekrutacji, ponieważ przypuszczalnie nie będą miały cech uchodzących w społeczeństwie za męskie (np. nie będą odpowiednio samodzielne, asertywne czy biegłe w ścisłych dziedzinach).

Czy wobec tego dyskryminacja statystyczna może zostać usankcjonowana swoją racjonalnością i widoczną optymalizacją w zarządzaniu ryzykiem? Sugeruję, że statystyczne uzasadnienie czy też ustalone w toku badań prawdopodobieństwo nie jest wystarczającym czynnikiem rozstrzygającym kwestię dopuszczalności tego typu różnicowania. Teoria relewantności statystycznej tej formy dyskryminacji koresponduje przy tym ze współczesnymi rozprawami z zakresu filozofii języka, które koncentrują się na generalizacjach[119]. Statystyka co prawda uwzględnia wiele czynników, wpływając na dokładniejszą i przez to lepszą predykcję, jednak dane, na których budowane jest przeświadczenie o probabilistycznym związku między wskaźnikami, oparte są na badaniach populacyjnych obarczonych ogromną luką informacyjną.

Problem ten został szeroko opisany przez Caroline Criado Perez na tle zaobserwowanego przez autorkę niedoboru danych dotyczących kobiet[120]. Już z tej przyczyny, mimo że dane są zasadniczo prawidłowe, ze względu na swoją niekompletność mogą one prowadzić do ułomności wnioskowania opartego na prawdopodobieństwie. Twierdzenia opisowe często nie są bowiem uniwersalnie prawdziwe, lecz nadal istnieją w powszechnej świadomości społecznej[121], w związku z czym również sposób prowadzenia badań i dokonywania analiz danych może mieć wpływ na prawidłowość wniosków.

Wreszcie informacje ustalone w procesie analizy statystycznej mogą być błędne (np. dane mogą wynikać z uprzedzeń i nie być poparte wiedzą naukową).

Należy przy tym wskazać, że niektóre statystycznie prawidłowe wskaźniki funkcjonują jako dowód wcześniejszego niesprawiedliwego różnicowania, ukazując historię dyskryminacji zapisaną w danych. Dlatego też formułowanie na ich podstawie wiążących konkluzji ograniczających prawa lub możliwości jednostki będzie oznaczać powrót do niechlubnej historii marginalizacji słabszych grup społecznych. Z pozoru neutralne porównawcze formuły matematyczne mogą przez to jedynie utrwalać niesłuszne i szkodliwe przekonania istniejące

[119] S.J. Leslie, *Generics Oversimplified*, „NOUS" 49, 2015, s. 28–54.
[120] C. Criado Perez, *op. cit*.
[121] Przykładowe generalizacje: ukąszenia komarów powodują choroby, samochody marki Volvo są niezawodne, matematycy są mądrzy, a ser szwajcarski ma dziury.

41:91571

w społeczeństwach, wspierając je dodatkowo motywacją statystyczną i rzekomą neutralnością danych. W związku z tym istotnym pytaniem jest to, w jaki sposób, w jakich obszarach i wobec kogo dopuścić statystyczne różnicowanie w procesie podejmowania decyzji.

Dyskryminacja statystyczna zwana jest również dyskryminacją racjonalną[122], gdyż niesie z sobą bezsprzecznie oczywiste korzyści płynące z optymalizacji. Z tego względu zwiększanie jej wpływu na stosunki społeczne wydaje się nieuniknione, co dokładnie przedstawię w rozdziale II niniejszej pracy. W tym miejscu nadmienię jedynie, że zakazanie możliwości kształtowania procesów decyzyjnych na podstawie statystyki wydaje się gospodarczo i strukturalnie nierealne.

W literaturze problem ograniczenia wpływu błędów statystycznych na różnicowanie poddaje się analizie z perspektywy możliwej indywidualizacji i randomizacji wyników[123]. Pierwsza z tych metod napotyka jednak na istotne trudności wdrożeniowe. Rzeczywista indywidualizacja jest bowiem kosztowna, a nadto organizacyjnie utrudniona. Narzucenie obowiązku indywidualizacji znosi więc wszelkie korzyści wynikające zarówno z optymalizacji zasobów, czasu i personelu, jak i zaangażowania środków finansowych w proces decyzyjny. Co więcej, na tej drodze natrafia się na prawne problemy związane z ochroną danych osobowych czy prywatności jednostki. Drugie rozwiązanie – randomizacja procesów – wydaje się pragmatyczne, a przez to bardziej prawdopodobne wdrożeniowo. Nie jest to tymczasem alternatywa pozbawiona wad i ryzyk, wynikających z mniejszej skuteczności predykcji oraz wyższych kosztów. W takim przypadku wskaźniki statystyczne powinny być korygowane o wartości graniczne lub losowe próby. Uważam też, że dyskryminacja statystyczna powinna być monitorowana pod względem jakości danych statystycznych, na których się opiera, oraz pod względem neutralności skutków różnicowania dla grup historycznie marginalizowanych. Oznaczałoby to, że decydent byłby zobowiązany weryfikować nie tylko prawidłowość wykorzystywanych danych statystycznych, ale również wyniki procesu decyzyjnego pod względem ich ewentualnego nieproporcjonalnego wpływu na marginalizację i wykluczenie społeczności w wyższym stopniu narażonych na efekt błędnego koła dyskryminacji. Takie uzupełnienie na poziomie podstawowego wykorzystania statystyki pozwoliłoby na racjonalne aplikowanie narzędzi matematycznych.

[122] O.H. Gandy, *Coming to Terms with Chance: Engaging Rational Discrimination and Cumulative Disadvantage*, Farnham 2009.
[123] B.E. Harcourt, *Against Prediction. Profiling, Policing, and Punishing in an Actuarial Age*, Chicago-London 2006.

42:23065

Rola dyskryminacji statystycznej i dotychczasowe względne przyzwolenie na jej stosowanie ulega obecnie istotnej zmianie. Postęp technologiczny doprowadził bowiem do stworzenia narzędzi i technologii umożliwiających generowanie dużych ilości danych statystycznych. W nowoczesnych systemach analitycznych kategoryzowanie i poszukiwanie zależności statystycznych nie będzie się jednak odbywać z pomocą ludzi, a przy użyciu systemów komputerowych. Tak uzyskane dane będą mogły być wykorzystywane jednocześnie w wielu dziedzinach i sferach życia obywateli. Różnicowanie oparte na statystyce staje się przy tym bardziej efektywne, co dodatkowo stanowi gospodarcze i polityczne uzasadnienie wyboru tej metody jako osiągającej najlepsze efekty najniższym kosztem. Jednocześnie wstępnie zarysowane zagrożenia wynikające ze statystycznej dyskryminacji mogą mieć istotny wpływ na grupy historycznie marginalizowane, które realnie dysponują mniejszymi środkami i zasobami do przeciwstawiania się dyskryminacji.

Trzeba przy tym stale pamiętać, że dyskryminacja jest fenomenem polityczno-społeczno-prawnym. Podsumowując rozważania rozdziału I, należy więc wskazać, że zjawisko różnicowania prowadzące do odpowiedzi prawa w postaci przepisów antydyskryminacyjnych jest ściśle powiązane z kontekstem zmian ustrojowych, społecznych i gospodarczych. Historyczne tło walki o równość wpływa bowiem na sposób kreowania definicji dyskryminacji. Wskazuje nam ponadto domniemanych oprawców – podmioty dominujące, posiadające społeczną władzę różnicowania – oraz typowe ofiary – tj. grupy uciskane, marginalizowane, stygmatyzowane i wyzyskiwane. W tym kontekście prawo było bazą, a nie nadbudową tych procesów różnicowania. Dopiero na skutek buntu społecznego władza wprowadzała zmiany, których zakres był ograniczany interesami grupy rządzącej. Tak wykształcone cechy warunkujące różnicowanie dyskryminujące są znamienne dla przedstawicieli tych grup, a przy tym umożliwiają różnicowanie jawne, publiczne i masowe.

Wykształcenie zakazu różnicowania właśnie ze względu na wskazane cechy reprezentatywne dla grup marginalizowanych ma uzasadnienie w postulatach równouprawnienia. Droga do równości wymaga jednak nie tylko uformowania się wspólnoty członków społeczności dyskryminowanych, ale również sprzeciwu poprzedzonego polityczną i grupową mobilizacją. Odpowiednie uświadomienie procesów towarzyszących postępowi technologicznemu może przy tym doprowadzić do efektywnego przeciwdziałania dezaktualizacji ochrony przed nowymi formami dyskryminacji, w tym przed jej statystycznym wymiarem, który w dobie masowej analizy danych będzie dynamicznie zyskiwał na znaczeniu.

43:63707

II. Sztuczna inteligencja i jej wpływ na rozwój automatycznych systemów decyzyjnych

Dyskryminacja ewoluuje wraz ze zmianami społeczno-kulturowymi, w związku z czym istotna jest eksplikacja pojęcia i treści prawa do niepodlegania krzywdzącemu różnicowaniu w dobie postępującej cyfryzacji. Celem tego rozdziału jest nakreślenie siatki pojęciowej obrazującej w sposób uproszczony charakterystykę działania automatycznych systemów decyzyjnych, w tym algorytmów sztucznej inteligencji. Przedstawienie treści i zakresu pojęć ma przy tym charakter wstępny do rozważań merytorycznych o prawnym problemie dyskryminacji przez automatyczne systemy decyzyjne (dalej: ASD) i jest niezbędne do zidentyfikowania zagrożeń związanych ze stosowaniem algorytmów analizujących duże zbiory danych w procesach decyzyjnych. W tym kontekście ukazuję problem różnicowania przez ASD i podaję jego przykłady prowadzące do niekorzystnych skutków społecznych.

Specyfika i mechanizmy leżące u podstaw uczenia maszynowego i analizy *big data* uświadamiają nam zmianę jakościową i ilościową zagrożeń, na które powinno zareagować prawo antydyskryminacyjne. Dodatkowo gdy zestawimy te niebezpieczeństwa ze schematem ukazanej w rozdziale I walki o równouprawnienie, uwidocznią się istotne różnice, które mogą utrudniać rozwój zakazu dyskryminacji zgodnie z historycznie wykształconymi wzorcami poszukiwania cech ludzkich identyfikujących grupy marginalizowane. W tym kontekście w ramach niniejszego rozdziału przedstawię relacje między automatycznymi systemami decyzyjnymi a procesami wykluczenia i dyskryminacji.

1. Automatyczne systemy decyzyjne

Podstawą normatywną zarówno ludzkich, jak i algorytmicznych systemów decyzyjnych jest zespół kryteriów związanych ze sposobem różnicowania, który określa oczekiwane zachowanie, priorytety i poziomy efektywności decyzji oraz kompetencje do realizowania władztwa w danym zakresie[124]. Ludzki proces decyzyjny to zatem droga wnioskowania, której celem jest rozwiązanie problemu. Zasadniczo proces ten w wypadku jego wykonania przez człowieka rozpoczyna się od ustalenia i analizy problemu do rozwiązania (określenia celu procesu decyzyjnego), po czym następuje gromadzenie faktów i danych relewantnych do podjęcia decyzji. Następnie zebrane informacje są wykorzystywane do formułowania

[124] E. Bayamlıoğlu, R. Leenes, *The "Rule of Law" Implications of Data-Driven Decision-Making: A Techno-Regulatory Perspective*, „Law, Innovation and Technology" 10, 2018, nr 2, s. 295–313.

44:42312

alternatywnych kierunków działania. Człowiek – jako decydent – szacuje nie tylko autonomicznie szanse związane z możliwym przebiegiem procesu, ale i maksymalny poziom ryzyka z punktu widzenia celu i zasobów leżących u podstaw procesu decyzyjnego. W dalszej kolejności, dysponując pozyskanymi informacjami, podejmuje decyzję, rozumianą jako wybór najlepszego rozwiązania problemu. To człowiek określa wówczas subiektywnie, jakie elementy czy kryteria składają się na najlepsze rozstrzygnięcie (tj. wyznacza priorytety, którymi mogą być koszty, szybkość, stopień dokładności analizy, pochodzenie informacji służących do różnicowania czy minimalizacja ryzyka).

Ludzki proces decyzyjny jest złożony, niekoherentny, a dodatkowo cechuje go zawodność. Na błędny wynik decyzji człowieka mogą mieć wpływ: niewystarczająca wiedza decydenta, jego doświadczenia osobiste i zawodowe, pośpiech, brak dostatecznej kontroli, indywidualne uprzedzenia, wartości moralne, utrwalone stereotypy, zwyczaje kulturowe, a nawet cele, historia i założenia organizacji, której interesy on reprezentuje. Na jakość decyzji może mieć wpływ również błędne określenie priorytetów, polegające na samodzielnym obraniu lub narzuceniu decydentowi nietrafnych kryteriów wyboru najlepszego rozwiązania, tj. zaakceptowanie wyższego ryzyka błędnej decyzji z uwagi na oszczędności w obszarze zasobów ludzkich lub finansowych. Szeroka elastyczność doboru kryteriów z jednej strony powoduje ryzyko nietrafnych decyzji i utrudnia kontrolę metodologii obranej przez decydenta, z drugiej zaś pozwala mu brać pod uwagę czynniki, które są specyficzne dla konkretnego przypadku lub sytuacji i które nie mogłyby być określone w sztywnych, odgórnie narzuconych regułach[125].

Zwiększanie roli informatyki w podejmowaniu decyzji dotyczących praw i obowiązków człowieka oddala proces decyzyjny od ogólnoludzkich standardów postępowania, przesuwając się w kierunku bardziej jednoznacznie określonych reguł, opartych na matematycznych modelach[126]. Automatyzacja procesu decyzyjnego może przy tym ograniczać ludzki wpływ na wszystkie lub niektóre etapy procesu decyzyjnego, umożliwiając decydentom operowanie na specjalistycznej wiedzy sprofilowanej przez hybrydowe systemy eksperckie.

W zależności od dziedziny, w której następuje automatyzacja, przykłady automatycznych systemów decyzyjnych[127] mogą być rozmaite. Skala postępu technologicznego w ostatnich

[125] R. Abebe *et al.*, *Roles for Computing in Social Change*, [w:] *FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, [b.m.w.] 2020, s. 254.
[126] *Ibidem*.
[127] Przepisy prawa posługują się pojęciem decyzji automatycznych, nie wprowadzając ich definicji legalnej. Można natomiast wskazać na pojęcie bliskoznaczne: „automatyczne przetwarzanie", zdefiniowane w art. 2 lit. c)

45:53644

dziesięcioleciach wpływa przy tym na to, że obecnie praktycznie każda branża i dziedzina życia znajduje się pod wpływem cyfryzacji[128]. Każdy z tych systemów ma na celu rozwiązanie pewnego problemu, a więc wypełnia zadania przypisywane procesowi decyzyjnemu. Automatyzacja może przy tym dotyczyć wyboru metody lub narzędzi do wykonania określonego działania w celu optymalizacji efektu, ale może się również odbywać na innych poziomach rozstrzygnięć. Mowa tu przede wszystkim o zbieraniu i analizie danych (ang. *text and data mining*)[129] potrzebnych do wyznaczenia alternatywnych kierunków działania, jak budowanie profilu decyzyjnego lub analiza profilu pod kątem podjęcia decyzji w indywidualnych sprawach[130].

Celem automatyzacji jest więc nie tylko usprawnienie technologiczne, lecz także lepsza jakość decyzji, rozumiana jako skuteczność predykcyjna, statystyczna trafność rozstrzygnięcia przy jednoczesnym podejmowaniu decyzji w wysoko wyspecjalizowanych lub bazujących na dużej ilości zmiennych kwestiach. W związku z tym nie każdy automatyczny system decyzyjny będzie w efekcie wydawał decyzje rozumiane jako rozstrzygnięcia o prawach i obowiązkach człowieka, jednak większość z nich będzie ostatecznie wpływała na sytuację ludzi. Decyzje algorytmiczne będą przy tym kreować ten status częściowo niedostrzegalnie i w sposób odbiegający od ludzkich schematów rozstrzygania o problemach.

Mechanizmem będącym „cyfrowy umysłem" ASD są matematyczne algorytmy. Przez pojęcie algorytmu można za Robin K. Hill rozumieć „konstrukt matematyczny ze skończoną, abstrakcyjną, efektywną, złożoną strukturą kontrolną, imperatywnie podaną, realizującą określony cel zgodnie z określonymi instrukcjami"[131]. Niespecyficzne rozumienie terminu

konwencji nr 108 Rady Europy z 28.01.1981 r. o ochronie osób w związku z automatycznym przetwarzaniem danych osobowych (Dz.U. z 2003 r. Nr 3, poz. 25), gdzie „automatyczne przetwarzanie" oznacza następujące operacje wykonane w całości lub częściowo za pomocą procedur zautomatyzowanych: gromadzenie danych, stosowanie do nich operacji logicznych lub arytmetycznych, ich modyfikowanie, usuwanie, wybieranie albo rozpowszechnianie.

[128] W medycynie mogą to być systemy wspierające lekarzy, pomagające analizować wyniki badań lub skuteczność metod leczenia; w przedsiębiorstwach wytwórczych – roboty zwiększające wydajność pracy i ograniczające liczbę błędów; w rolnictwie – narzędzia do automatycznego zbierania plonów czy regulowania nawodnienia upraw; w ubezpieczeniach – algorytmy obliczające ryzyko zaistnienia wypadku ubezpieczeniowego w celu określenia wysokości składek.

[129] Zgłębianie danych (ang. *data mining*) to proces analityczny przeznaczony do badania dużych zasobów danych (zazwyczaj powiązanych z zagadnieniami gospodarczymi lub rynkowymi) w poszukiwaniu regularnych wzorców oraz systematycznych współzależności pomiędzy zmiennymi, a następnie służący do oceny wyników poprzez zastosowanie wykrytych wzorców do nowych podzbiorów danych. Ostatecznym celem *data mining* jest najczęściej przewidywanie (zachowań klientów, wielkości sprzedaży, prawdopodobieństwa utraty klienta itp.); por. E. Bayamlıoğlu, R. Leenes, *op. cit.*

[130] *Ibidem*.

[131] R.K. Hill, *What an Algorithm Is*, „Philosophy & Technology" 29, 2016, nr 1, s. 47; jeśli nie podano inaczej, z angielskiego przeł. M.D.

46:20745

„algorytm" dotyczy również technologii komputerowej skonfigurowanej do wykonania konkretnego zadania[132]. Takie ukształtowanie pojęcia mogłoby wypełniać treść podstawy normatywnej systemu decyzyjnego, przy czym opracowywanie instrukcji algorytmu (tj. ustalenie sekwencji operacji) oraz dobór danych i ich wagi co do zasady pozostają aktualnie domeną ludzką.

Wraz z postępem cyfryzacji jednak i ta kompetencja w coraz mniejszym zakresie podlega władztwu i realnemu nadzorowi człowieka. Tempo rozwoju informatyki pozwala przypuszczać, że w niedalekiej przyszłości systemy automatyczne samodzielnie będą w stanie odnaleźć problem do rozwiązania i źródła informacji konieczne do przeprowadzenia analizy. Już teraz ASD wykorzystujące zaawansowane algorytmy zdolne są do wyprowadzania autonomicznych wniosków na podstawie dostępnych informacji. Uczące modele zawierają zaś jedynie listę instrukcji do wykonania na podstawie analizy przekazanych im danych uczących[133]. Należy przy tym zaznaczyć, że aktualnie większość ASD opiera się na algorytmach heurystycznych i jest wykorzystywana do rozwiązywania prostszych problemów optymalizacyjnych. Tego typu algorytmy nie uczą się na podstawie danych. Są zaprojektowane w sposób obejmujący konkretne strategie lub reguły, które mają być stosowane do rozwiązywania problemów[134]. W przeciwieństwie do tego sieci neuronowe uczą się na podstawie danych i mogą wykazywać zdolność do generalizacji na nowe dane. Do procesu uczenia potrzebują jednak dużej ilości danych treningowych, aby dostosować swoje wagi i parametry do rozwiązywania konkretnych problemów. Każdy neuron w takich sieciach jest połączony z innymi neuronami za pomocą wag określających siłę połączenia między neuronami i dostosowywanych podczas procesu uczenia. W związku z tym reguły leżące u podstaw procesów decyzyjnych są niepewne i częściowo nieprzewidywalne w zakresie rezultatów działania[135].

Emre Bayamlıoğlu i Roland Leenes, opisując proces decyzyjny ASD, wyjaśniają, że:

---

[132] B. Mittelstadt *et al.*, *The Ethics of Algorithms: Mapping the Debate*, „Big Data & Society. In press" 2, 2016, s. 1–21.

[133] Sieci neuronowe są często stosowane w zadaniach związanych z uczeniem maszynowym, takich jak rozpoznawanie obrazów, analiza tekstu, przewidywanie cen akcji, samochody autonomiczne i wiele innych. Są one bowiem zdolne do radzenia sobie z zadaniami, które wymagają rozpoznawania wzorców w danych.

[134] Algorytmy heurystyczne są nierzadko wykorzystywane przy problemach optymalizacyjnych, takich jak planowanie tras, szeregowanie zadań, optymalizacja produkcji, projektowanie układów elektronicznych i innych. Są używane tam, gdzie problem polega na znalezieniu najlepszego rozwiązania spośród wielu możliwości.

[135] M. Rojszczak, *Ochrona prywatności w cyberprzestrzeni z uwzględnieniem zagrożeń wynikających z nowych technik przetwarzania informacji*, Warszawa 2019.

47:65681

> w automatycznych systemach decyzyjnych opartych na danych reguły decyzyjne są (częściowo) dynamiczne, a normy narzucane przez te systemy nie są stabilne, lecz są przedmiotem trwałej i nieustannej rekonfiguracji; to zaś powoduje, że sama reguła decyzyjna wyłania się (autonomicznie) z danych (dynamicznych) wykorzystywanych do uczenia systemu[136].

Należy zgodzić się z autorami, że taka specyfika działania ASD prowadzi do nieprzejrzystości normatywnej automatycznych systemów decyzyjnych, gdyż zasady ich działania mogą być stale modyfikowane i dostosowywane do zmiennych. Ta zmienność i nieprzejrzystość połączona z nieprzewidywalnością rezultatów działania powoduje, że ASD wymagają nowej logiki regulacyjnej oraz podejścia opartego na inherentnym ryzyku błędu takich decyzji. Wniosek ten można również oprzeć na braku wrażliwości systemów na ludzką świadomość historii marginalizacji i kontekstów dominacji, które ją poprzedzały[137].

Wspólnym mianownikiem automatyzacji jest ograniczenie zaangażowania człowieka w proces (zarówno w aspekcie fizycznym, jak i umysłowym) przy jednoczesnym optymalizowaniu jego efektów. Algorytmy pozwalają bowiem ograniczyć błędy związane z niewłaściwą oceną statystycznego ryzyka podjęcia nietrafionej decyzji, gdyż ich możliwości analityczne w niekwestionowany sposób przekraczają ludzkie. Na poprawę jakości decyzji wydawanych przez ASD wpływa zatem możliwość przetworzenia wielkiej ilości danych, co następuje z dużą prędkością, wydajnością i dokładnością, jak również daje szansę poszukiwania w tych danych wzorców[138]. Przyspieszenie, usprawnienie oraz poprawianie trafności podjętych decyzji powinno jednak zostać zestawione i porównane ze skalą możliwego wpływu ASD na prawa i wolności człowieka[139]. Ponieważ wnioskowanie statystyczne lub techniki uczenia maszynowego wytwarzają prawdopodobną, ale niepewną wiedzę, statystyka zamiast rozumu wkracza w sferę wykonywania władztwa publicznego[140]. Potencjał predykcyjny[141] automatycznych systemów decyzyjnych jest niepodważalną zachętą do ich stosowania w sektorze prywatnym i publicznym, gdzie zarządzanie zasobami poddane jest dużej presji racjonalizacji[142]. Predykcja jest jednak ryzykowna, gdyż mimo odnalezienia korelacji pomiędzy wskaźnikami wnioski oceniane z dalszej perspektywy nadal mogą być błędne (np.

[136] E. Bayamlıoğlu, R. Leenes, *op. cit.*
[137] R. Abebe *et. al.*, *op. cit.*, s. 254.
[138] D. Morondo Taramundi, *op. cit.*, s. 74.
[139] B. Mittelstadt *et. al.*, *op. cit.*, s. 7.
[140] E. Bayamlıoğlu, R. Leenes, *op. cit.*
[141] Potencjał predykcyjny (ang. *predictive analytics*, zamiennie stosowane z *data mining* lub *data science*) rozumiany jest jako proces wydobywania informacji z istniejących zbiorów danych w celu określenia wzorów i przewidywania przyszłych zdarzeń i trendów.
[142] D. Freeman *et al.*, *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies*, Public Law Research Paper No. 20-54, NYU School of Law, New York 2020.

48:18077

będą dotyczyć jedynie konkretnego fragmentu społeczeństwa, o którym informacje są zapisane w analizowanych danych). Jako że grupy marginalizowane cechowało historyczne wykluczenie, objawiające się również w pomijaniu ich lub zakrzywianiu ich obrazu w opiniach społecznych, predykcja ASD może istotnie przyczynić się do wzrostu procesów dyskryminacji. Z tej przyczyny należy poświęcić nadzwyczajną uwagę stypizowaniu niebezpieczeństw pogłębiania przez ASD schematów wykluczenia. Stosowanie automatycznych systemów decyzyjnych może bowiem, podobnie do prawa utrwalającego stosunki dominacji, umacniać i konserwować niesprawiedliwość wynikającą z nierówności, a przy tym być dla tych procesów siłą napędową umożliwiającą ich przyspieszenie w wymiarze globalnym i masowym.

Decyzje podejmowane przez ASD mogą obecnie dotyczyć każdego, w każdym rodzaju relacji prawnej i w każdym momencie, bez względu na to, czy automatyzacja jest dostrzegalna przez adresata, czy też nie. Procesowi temu towarzyszy klasyfikowanie ludzi na podstawie danych, w tym ich śladów cyfrowych zostawianych w Internecie oraz pozostawianych przez użytkowane przez nich urządzenia cyfrowe. Mechanizmy te nie tylko różnicują na podstawie udostępnianych świadomie informacji, ale również mają potencjał tworzenia nowej wiedzy na skutek przekształcania i łączenia fragmentów danych we wcześniej nieistniejący zbiór. Dane te daleko wykraczają poza cechy chronione prawem antydyskryminacyjnym, a także ochroną danych osobowych czy prawem do prywatności. Analizowane systemy mogą przez to osiągać przewagę nad ludźmi i wpływać podświadomie na ich zachowania, przekształcając adresatów decyzji (w tym na skutek profilowania) w wydajnych kupujących, wyborców czy pracowników[143].

Przyjmuje się, że waga analizy dużych zbiorów danych pochodzących z różnych źródeł (np. Internet, sieci komunikacyjne, systemy transportowe, bankowe, ubezpieczeniowe, sieci energetyczne, jak również państwowe bazy danych) będzie rosła, a ich wykorzystywanie do poprawy jakości życia, działania państwa i optymalizacji konsumpcji nie zmniejszy się. Rozwiązania bazujące na automatycznych systemach decyzyjnych mają potencjał lepszego prognozowania, dzięki czemu mogą skuteczniej przydzielać zasoby oraz personalizować rodzaje świadczeń społecznych. Zakres zastosowania narzędzi informacyjno-

[143] C. O'Neil, *op. cit.*, s. 81. W książce autorka podała przykład Corinthian College, w wypadku którego prokurator ustalił, że uczelnia kierowała swoje reklamy do osób „osamotnionych", „niecierpliwych", „z niską samooceną", które miały wokół siebie „niewiele osób, którym na nich zależy" oraz które „utknęły" i nie są w stanie „przewidywać i odpowiednio planować swojej przyszłości". W skardze praktyki stosowane przez Corinthian College uznano za „bezprawne, niesprawiedliwe i oszukańcze" – *ibidem*, s. 109.

49:76746

komunikacyjnych doprowadził już do pojawienia się nowego pojęcia – „cyfrowego *welfare state*"[144].

Wraz z rozpowszechnianiem się takich praktyk na poziomie publicznoprawnym pojawiają się pytania o sposób wykorzystania, legitymację do zbierania i analizy oraz konkurencję między postulatami cyfryzacji a prawami konstytucyjnymi. Istotna jest też granica uprawnienia władzy do korelowania danych, którymi dysponuje, z obserwacjami uzyskanymi dzięki analityce *big data*[145]. Pytania te pozostają aktualne również na płaszczyźnie prywatnoprawnej, albowiem decyzje wydawane przez automatyczne systemy decyzyjne mogą mieć zasadnicze znaczenie w wielu aspektach życia prywatnego, społecznego i zawodowego człowieka[146].

W niniejszym podrozdziale przedstawiam dwa przykłady wdrożenia ASD w obszarze wykonywania władztwa publicznego. W dalszych rozdziałach pracy będą one przedmiotem rozważań prawnych związanych z zagadnieniami automatyzacji procesów decyzyjnych w kontekście ryzyka wykluczenia i dyskryminacji.

Pierwszy przykład dotyczy polskiego systemu opartego na gromadzeniu danych i profilowaniu osób bezrobotnych, wdrożonego w maju 2014 r. przez Ministerstwo Pracy i Polityki Społecznej. Koncepcja metodologii profilowania ludzi była tutaj posadowiona na kategoryzacji bezrobotnych w celu ustalenia stopnia ich oddalenia od rynku pracy. W tym zakresie stworzono trzy grupy (tzw. profile), różniące się stopniem gotowości do podjęcia pracy i rodzajem pomocy, jaką bezrobotni otrzymają z urzędu pracy (np. pośrednictwo pracy, szkolenia zawodowe, staże, dodatek aktywizacyjny). W ASD przyjęto następujący podział analizowanych osób:

1. poszukujący pracy z odpowiednimi kwalifikacjami zawodowymi i umiejętnościami interpersonalnymi;
2. poszukujących pracy „wymagający wsparcia", ponieważ ich umiejętności zawodowe są ograniczone lub przestarzałe;

[144] Stosuje je np. Organizacja Narodów Zjednoczonych w raporcie *Cyfrowe państwa opiekuńcze i prawa człowieka – Raport specjalnego sprawozdawcy ds. skrajnego ubóstwa i praw człowieka*, nr A/74/493 z 11.10.2019 r.

[145] J. Niklas, K. Sztandar-Sztanderska, K. Szymielewicz, *Profiling the Unemployed in Poland: Social and Political Implications of Algorithmic Decision Making*, Warszawa 2015, https://panoptykon.org/sites/default/files/leadimage-biblioteka/panoptykon_profiling_report_final.pdf (dostęp: 4.12.2021).

[146] Racjonalność ekonomiczna jest oparta na teorii decyzji, która łączy logikę i teorię prawdopodobieństwa z teorią użyteczności oraz umożliwia podejmowanie decyzji w warunkach niepewności; por. M.N. Huhns, *Being and Acting Rational [Agent Design]*, „IEEE Internet Computing" 7, 2003, nr 2, s. 92.

50:10145

3. osoby poszukujące pracy, które są „dalekie od rynku pracy".

Na skutek takiej kategoryzacji w nieproporcjonalnym stopniu do trzeciej grupy zaliczono kobiety, osoby starsze, osoby z niepełnosprawnościami oraz osoby z niewielkim wykształceniem lub bez formalnego wykształcenia[147]. Niewyrażenie zgody na ustalenie profilu pomocy zgodnie ustawą[148] skutkowało wydaniem przez właściwego starostę decyzji o pozbawieniu osoby bezrobotnej statusu bezrobotnego, a zatem świadczeń pieniężnych i socjalnych. Bezrobotni zarówno nie mieli dostępu do zasad działania algorytmu, jak i nie znali kryteriów prowadzących do sprofilowania ich w ramach konkretnej grupy. Decyzja automatyczna mogła przy tym zostać zmieniona przez urzędnika korzystającego z systemu, lecz nie mogła być zakwestionowana przez bezrobotnego. Mimo istniejącego uprawnienia decydenci nie korzystali jednak z prawa zakwestionowania decyzji algorytmicznej, nawet gdy jej treść wydawała się oczywiście wątpliwa[149].

System stworzony i wdrożony przez polskie ministerstwo był przedmiotem krytyki zarówno przedstawicieli nauk prawnych, jak i organizacji pozarządowych zajmujących się badaniem problemu dyskryminacji (m.in. Fundacji Panoptykon[150]). Ostatecznie ustawa powołująca system została zakwestionowana przez Rzecznika Praw Obywatelskich Adama Bodnara i podana do Trybunału Konstytucyjnego[151]. Na skutek pozytywnego rozpoznania wniosku o stwierdzenie niezgodności automatycznego systemu z prawem TK częściowo orzekł o utracie mocy obowiązującej kwestionowanej ustawy, pozostawiając ją jednak przez 12 miesięcy w obrocie[152].

Kolejnym przykładem wykorzystania automatycznych systemów decyzyjnych w sferze publicznoprawnej była sprawa niderlandzkiego Systeem Risicoindicatie (dalej: SyRI)[153], będącego narzędziem algorytmicznym stosowanym przez organy państwowe Królestwa Niderlandów w celu zapobiegania nadużyciom i ich zwalczania w dziedzinie ubezpieczeń społecznych i systemów świadczeń społecznych, opierającym się na analizie informacji o

[147] Fundacja Panoptykon stwierdziła, że rząd zapewnił skąpe zasoby w kwestii zatrudnienia osobom poszukującym pracy w trzeciej kategorii, odcinając je od krytycznych środków wspierania zatrudnienia, takich jak staże finansowane ze środków publicznych, zwrot kosztów opieki nad dziećmi i szkolenia; por. J. Niklas, K. Sztandar-Sztanderska, K. Szymielewicz, *op. cit*.
[148] Ustawa z 20.04.2004 r. o promocji zatrudnienia i instytucjach rynku pracy, Dz.U. z 2016 r. poz. 645 ze zm.
[149] Pracownicy polskich Powiatowych Urzędów Pracy akceptowali sugestie dyskryminującego systemu w 99,4% przypadków; por. J. Niklas, K. Sztandar-Sztanderska, K. Szymielewicz, *op. cit*.
[150] *Ibidem*.
[151] Wniosek Rzecznika Praw Obywatelskich, nr VII.501.277.2014.KM.
[152] Wyrok TK z 6.06.2018 r., sygn. K 53/16.
[153] Sprawa nr C-09-550982-HA ZA 18-388.

51:82517

wysokości dochodów, danych z ubezpieczeń społecznych oraz zatrudnienia[154]. W rzeczywistości był to algorytm przetwarzający informacje o obywatelach i rezydentach, które zostały zgromadzone w krajowych rejestrach zatrudnienia, systemach świadczeń socjalnych, raportach instytucji kredytowych oraz historii edukacji i zameldowania. Na podstawie tych danych system kategoryzował obywateli w profilach ryzyka. Kiedy SyRI określił daną osobę jako potencjalnego oszusta, powiadamiał odpowiednią jednostkę rządową, która była zobligowana do wszczęcia dochodzenia w sprawie możliwego zachowania przestępczego lub czynów niezgodnych z prawem w terminie dwóch lat od zawiadomienia.

System był wykorzystywany do badania dzielnic, które zostały uznane za obszary statystycznie bardziej narażone na nadużycia (o wyższych współczynnikach przestępczości), w związku z czym wywołał powszechną krytykę związaną z zarzutem dyskryminacji osób ubogich i mniejszości narodowych. Osoba zakwalifikowana do kategorii ryzyka nie mogła przy tym ustalić, na podstawie jakich danych została uznana za potencjalnego przestępcę, nie mogła również zakwestionować tej predykcji.

Ostatecznie 5.02.2020 r. niderlandzki sąd z siedzibą w Hadze nakazał wstrzymanie stosowania SyRI[155]. Sąd uzasadnił swoją decyzję tym, że organ administracji, który częściowo opiera swoje działania na automatycznym systemie decyzyjnym, nie jest w stanie właściwie uzasadnić przyczyn i motywacji swojego rozstrzygnięcia. Sąd uznał również, że algorytm SyRI ze względu na brak przejrzystości naruszał prawo do poszanowania życia prywatnego i rodzinnego.

Powszechne stosowanie systemów opartych na danych, w tym ich wykorzystanie w procesach decyzyjnych realizowanych przez władzę publiczną, wskazuje na naszą obecną i przyszłą zależność od technologii upowszechniających i wzmacniających racjonalność obliczeniową, łączącą cele ze środkami w nowatorski i niezrozumiały dla człowieka sposób[156]. Zwiększenie probabilistycznej prawidłowości takich decyzji przy jednoczesnym ograniczeniu kosztów procesu decyzyjnego i braku efektywnego nadzoru weryfikującego na bieżąco błędy może wpływać na nieuzasadnione przecenianie jakości tego typu rozstrzygnięć.

[154] Opieranie roszczeń bezpośrednio na międzynarodowych zobowiązaniach w zakresie praw człowieka, a szczególnie na EKPCz zamiast na niderlandzkiej konstytucji, jest powszechną praktyką prawną w Niderlandach. Wynika to z przepisów konstytucyjnych, które zabraniają sądom dokonywania kontroli konstytucyjnej przepisów prawnych, ale pozwalają na bezpośrednie stosowanie międzynarodowych traktatów dotyczących praw człowieka.
[155] Sprawa nr C-09-550982-HA ZA 18-388.
[156] E. Bayamlıoğlu, R. Leenes, *op. cit*.

52:38849

Komisja Europejska opublikowała już uwagę, że wyniki uzyskane z ASD mają pozornie obiektywny i niepodważalny charakter, do którego ludzki decydent może przywiązywać zbyt dużą wagę, zrzucając tym samym z siebie odpowiedzialność za prawidłowość i zgodność z prawem danej decyzji[157]. Nadzór ludzki nad ASD może mieć więc charakter pozorny lub niewystarczający. Potwierdza to przypadek polskiego ASD do profilowania bezrobotnych, którego operatorzy ulegali tzw. tendencji do automatyzacji. Dalsze ryzyko zrealizowało się zaś w sytuacji SyRI, w wypadku którego ASD miał przewagę informacyjną nad operatorem, a nadto operator systemu nie mógł go kontrolować w czasie rzeczywistym ze względu na szybkość przetwarzania i mnogość operacji.

Delegacja wykonywania władzy publicznej nie tylko prowadzi do intensyfikacji zagrożeń związanych z samą nieprzejrzystością epistemiczną omawianych systemów, przez co kognitywny agent nie jest w stanie poznać uzasadnienia i podstaw decyzji, ale w przypadku bezrefleksyjnego zawierzenia statystyce grozi również wpisaniem w decyzje błędów i uprzedzeń istniejących w analizowanych danych populacyjnych[158]. System automatyczny może zatem działać poprawnie, wprowadzając nowe definicje sprawiedliwości algorytmicznej w zakresie uśrednień i prognoz, jednak mimo tego wyniki nie będą przystawać do reprezentowanych w nowożytnych systemach prawnych zespołów wartości i gwarancji obywatelskich. To zaś wymaga przeprowadzenia analizy prawnej wpływu różnicowania przez takie systemy na postulaty równościowe i problem dyskryminacji. Kwalifikacja i ocena prawna muszą jednak zostać poprzedzone poznaniem przedmiotu badania, a więc elementów składowych mających wpływ na wynik automatycznego procesu decyzyjnego.

2. Pojęcie sztucznej inteligencji, uczenia maszynowego i *big data*

W niniejszym podrozdziale przedstawię siatkę pojęciową związaną z automatycznymi systemami decyzyjnymi, w tym terminy wykorzystywane w ramach analizy mechanizmów leżących u podstaw obserwowanego postępu technologicznego, tj. słabą i silną sztuczną inteligencję, głębokie uczenie maszynowe, sieci neuronowe, *big data*, a także wyjaśnię specyfikę analizy dużych zbiorów danych. Zrozumienie działania ASD jest bowiem kluczowe

[157] Zmieniony wniosek dotyczący dyrektywy Rady w sprawie ochrony osób fizycznych w zakresie przetwarzania danych osobowych i swobodnego przepływu tych danych, COM(92) 422 final – SYN 287.
[158] P. Humphreys, *The Philosophical Novelty of Computer Simulation Methods*, „Synthese" 169, 2009, nr 3, s. 615–626.

53:70383

do oceny istniejących i opracowywanych rozwiązań prawnych mających ograniczyć niepożądane konsekwencje stosowania tego typu systemów z perspektywy adekwatności przepisów prawa oraz narzędzi regulacyjnych wpisywanych w nowe akty. Jako że obowiązujące prawo antydyskryminacyjne wykształciło się w odpowiedzi na procesy zmierzające do ograniczenia niesprawiedliwego różnicowania zauważalnego w historycznie specyficznych formach, zmiana tych ostatnich może wpłynąć na dezaktualizację instrumentów prawa.

Najbardziej zaawansowanym systemem automatycznego podejmowania decyzji jest sztuczna inteligencja (ang. *artificial intelligence*, niem. *künstliche Intelligenz*; dalej: SI). Pojęcie to funkcjonuje zarówno w języku potocznym, literackim, jak i naukowym, przy czym nie ma uniwersalnej definicji umożliwiającej jednolite uchwycenie jego przedmiotu. Ta nieprecyzyjność pojęciowa łączona jest z nowością dziedziny oraz jej ciągłą ewolucją. Kwestia pojęciowa nie jest przy tym mankamentem dziedziny, albowiem – jak ukazują aktualnie trwające prace legislacyjne – istnienie definicji SI może zawężać przedmiot badań lub zakres i funkcję regulacji[159]. Ze względu na wykreowanie swoistej mody na SI, ale i popyt na produkty czy usługi na niej oparte wiele podmiotów wykorzystuje hasła sztucznej inteligencji w sposób mylący. Warto dodać, że sztuczna inteligencja powszechnie kojarzy się z humanoidalnymi robotami oraz testem Turinga[160], a taka wizualizacja SI wydaje się niezasadnie odwracać uwagę od nagminności problemów ze stosowaniem ASD w życiu codziennym. Na potrzeby dalszych rozważań sugeruję więc jako sztuczną inteligencję definiować automatyczny system decyzyjny, który pozwala na wykonywanie zadań wymagających procesu uczenia się i uwzględniania nowych okoliczności w toku rozwiązywania danego problemu i który może w różnym stopniu – w zależności od konfiguracji – działać autonomicznie oraz wchodzić w interakcję z otoczeniem[161], a także podejmować na bazie analizy danych racjonalne decyzje, które w niektórych przypadkach może również cechować nieprzewidywalność[162]. Problem

[159] A. Michalak, *Odpowiedzialność cywilnoprawna w obrocie oprogramowaniem komputerowym w erze sztucznej inteligencji*, Warszawa 2021.
[160] W 1950 r. Alan Turing zaproponował test w ramach badań nad stworzeniem sztucznej inteligencji. Zadanie polegało na tym, że sędzia będący człowiekiem prowadził rozmowę w języku naturalnym z pozostałymi uczestnikami doświadczenia. Jeśli sędzia nie był w stanie wiarygodnie określić, czy któraś ze stron jest maszyną czy człowiekiem, wtedy mówi się, że maszyna przeszła test. Zakłada się, że zarówno człowiek, jak i maszyna próbują zdać test, zachowując się w sposób możliwie zbliżony do ludzkiego.
[161] T. Zalewski, *Definicja sztucznej inteligencji*, [w:] *Prawo sztucznej inteligencji*, red. L. Lai, M. Świerczyński, Warszawa 2020, rozdz. 3.
[162] A. Michalak, *op. cit.*

54:41237

kompleksowego uchwycenia językowej definicji SI jest jednak złożony, a rozwiązanie sporów doktrynalnych wykracza poza ramy niniejszej analizy.

W ramach systemów sztucznej inteligencji, a więc zaawansowanych automatycznych systemów decyzyjnych stworzonych do realizacji określonego celu, zdolnych do podejmowania samodzielnych decyzji, wyróżnia się słabą oraz silną SI[163]. Aleksander Chłopecki w książce poświęconej sztucznej inteligencji definiuje słabą SI jako system, który w interakcjach w obrocie gospodarczym mimo terminologicznej „słabości" działa w sposób samodzielny, tzn. ma wbudowane algorytmy samouczące się, powodujące, że działanie nie podlega ograniczonej (i z reguły następczej) kontroli osób fizycznych[164]. Natomiast silną sztuczną inteligencję dodatkowo cechuje samoświadomość, co oznacza zdolność do samostanowienia[165] i autonomiczne zdolności poznawcze[166]. Taki system może mieć predyspozycje do myślenia porównywalnego z ludzkim, czyli myślenia w sposób niesymulowany[167]. W przeciwieństwie do silnej SI w przypadku słabej SI sposób działania nie ma charakteru autonomicznego[168]. Przy takim założeniu silna sztuczna inteligencja byłaby zapewne obdarzona świadomością swego istnienia, przez co zyskiwałaby podmiotowość.

Rozróżnienie pomiędzy tymi systemami ma znaczenie dla dalszej analizy. W przypadku delegacji procesów decyzyjnych do systemów automatycznych mających status słabej sztucznej inteligencji zarówno rola człowieka, tworzone przez niego profile i instrukcje, jak i ferowany przez niego dobór i metoda analizy danych będą bowiem miały potencjalne znaczenie dla ograniczania ryzyka. W wypadku zaś silnej sztucznej inteligencji kompetencje i nadzór człowieka byłyby oczywiście niewystarczające, nie tylko wobec dysproporcji potencjału analitycznego systemu. Co ważne, aktualnie dostępne systemy SI to przykłady słabej sztucznej inteligencji.

Od sztucznej inteligencji należy odróżnić Internet rzeczy (ang. *Internet of Things*; dalej: IoT), który może operować na algorytmach sztucznej inteligencji. System ten został wprowadzony do obrotu w celu nadania komputerom umiejętności gromadzenia informacji bez udziału

[163] M. Rojszczak, *Prawne aspekty systemów sztucznej inteligencji – zarys problemu,* [w:] *Sztuczna inteligencja,* blockchain*, cyberbezpieczeństwo oraz dane osobowe. Zagadnienia wybrane,* red. K. Flaga-Gieruszyńska, J. Gołaczyński, D. Szostek, Warszawa 2019, Legalis/el., s. 4.
[164] A. Chłopecki, *Sztuczna inteligencja – szkice prawnicze i futurologiczne,* Warszawa 2021.
[165] M. Rojszczak, *Prawne aspekty systemów*…, s. 4.
[166] A. Chłopecki, *op. cit*.
[167] T. Zalewski, *op. cit*.
[168] A. Krasuski, *Status prawny sztucznego agenta. Podstawy prawne zastosowania sztucznej inteligencji,* Warszawa 2020.

55:11101

ludzkiego[169]. Urządzenia funkcjonujące w ramach IoT zarówno zbierają dane z otoczenia oraz te wprowadzane przez człowieka, jak i pozyskują informacje dzięki wbudowanemu systemowi radiowej komunikacji[170]. Internet rzeczy jest przy tym zdolny do wyczuwania, uruchamiania i komunikacji między sobą i ze środowiskiem urządzeń zdolnych do dzielenia się informacją i oddziaływania przynajmniej w części autonomicznie na świat fizyczny (ang. *smart objects*). Co istotne, IoT spełnia cel założony przez jego pomysłodawców, generując ogromne ilości danych, które mogą być pozyskiwane w czasie rzeczywistym lub niemalże rzeczywistym. Dzięki temu informatycy zyskali narzędzie do budowania baz danych o ludziach bez ich udziału, a nawet zezwolenia. Tak zgromadzone informacje mogą obrazować schematy i przyzwyczajenia danego fragmentu populacji, nadając im charakter normy, jednocześnie pozostawiając znaczącą lukę w danych dotyczących pozostałej części społeczeństwa, których systemy nie gromadzą (często chodzi o informacje o odpowiednio uboższej czy z innych względów niedoreprezentowanej grupie lub której dane z przyczyn podwyższonych standardów ochrony prywatności są trudniej dostępne, jak informacje dotyczące ludzi najbardziej zamożnych).

Obserwowany w ostatnim okresie dynamiczny rozwój sztucznej inteligencji wynika z udoskonalonych rozwiązań uczenia maszynowego, zwiększenia mocy obliczeniowej współczesnych komputerów i dostępności dużych zbiorów danych potrzebnych do trenowania algorytmów. Systemy SI w celu udoskonalania działania mogą też wykorzystywać symboliczne reguły[171] albo uczyć się modelu numerycznego, a także zmieniać swoje działanie na podstawie analizy wpływu ich poprzednich decyzji na otoczenie[172].

[169] E.M. Kwiatkowska, *Rozwój Internetu rzeczy – szanse i zagrożenia*, „Internetowy Kwartalnik Antymonopolowy i Regulacyjny" 8, 2014, nr 3. s. 61.

[170] RFID (ang. *radio frequency identification*) to technologia pozwalająca na przechwytywanie danych dzięki wykorzystaniu znaczników reagujących na sygnał radiowy lub emitujących sygnały radiowe, które są odczytywane przez czytniki, w celu umożliwienia śledzenia przemieszczania się różnych sprzętów, a także ich identyfikacji. Technologia jest wykorzystywana w systemach automatycznego rozpoznawania rzeczy.

[171] Metoda ta zakłada, że zasady tworzone są w drodze ludzkiej interwencji (zob. Biała Księga). Opiera się ona na wykorzystaniu słów lub symboli zrozumiałych dla człowieka, uzupełnionych o reguły, dzięki którym program może połączyć i przetworzyć te symbole w celu osiągnięcia zamierzonego celu; por. A. Michalak, *op. cit*.

[172] Niezależna grupa ekspertów wysokiego szczebla ds. sztucznej inteligencji powołana przez Komisję Europejską w czerwcu 2018 r., *Wytyczne w zakresie etyki dotyczące godnej zaufania sztucznej inteligencji*, Komisja Europejska, Bruksela 2019, s. 48, https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/JURI/DV/2019/11-06/Ethics-guidelines-AI_PL.pdf (dostęp: 4.12.2021).

56:10934

Uczenie się systemu może mieć charakter nadzorowany[173] lub przebiegać autonomicznie[174]. Uczenie z nadzorem dotyczy rozwoju modeli obliczeniowych dla prognozowania wyników opartych na historycznych danych wejściowych i wyjściowych[175]. Z kolei w przypadku uczenia się bez nadzoru modele obliczeniowe budowane są wyłącznie na podstawie istniejących danych wejściowych, przy czym brakuje powiązanych danych wyjściowych, które mogłyby zostać wykorzystane w celach treningowych[176]. Algorytm oparty na nienadzorowanym uczeniu maszynowym umożliwia zatem rozpoznawanie obrazu czy analizę tekstu, jednak dane wyjściowe systemu mogą być nieprzewidywalne lub nieintuicyjne.

W tym kontekście należy wyjaśnić jeszcze trzy pojęcia związane z uczeniem się algorytmów, mianowicie: uczenie maszynowe, głębokie uczenie maszynowe i sieci neuronowe. Pierwsze z nich – uczenie maszynowe – to termin z zakresu sztucznej inteligencji o szerokim znaczeniu[177]. W przeciwieństwie do tradycyjnych form analizy danych, których celem jest opracowanie zestawień lub statystyk w odpowiedzi na konkretne zapytanie, uczenie maszynowe jest zbiorem technik poszukiwania związków statystycznych umożliwiających komponowanie nowych wzorców i wytwarzanie nowej wiedzy oraz generowanie modeli dających się wykorzystać do skutecznych przewidywań (predykcji)[178]. Dane te mogą być przedmiotem oceny jako wcześniej opracowany zbiór lub analiza może obejmować ich strumień w czasie rzeczywistym[179].

Drugi termin – głębokie uczenie maszynowe – odnosi się do zaawansowanej technologii uczenia algorytmów. Rozwiązanie to umożliwia modelowanie abstrakcji wysokiego poziomu danych za pomocą wielu warstw ich przetwarzania[180]. Może być ono realizowane na bazie

---

[173] Uczenie nadzorowane (ang. *supervised learning*) – jego model, mówiąc w uproszczeniu, opiera się na wyznaczeniu pewnych wzorców na podstawie konkretnych modelowych przykładów (zarówno pozytywnych, jak i negatywnych), a następnie weryfikowaniu operacji systemu, opierając się na tak wyodrębnionych modelowych przykładach; por. A. Michalak, *op. cit*.

[174] W technologii agentów inteligentnych (ang. *Intelligent Agent Technology*) „inteligentny agent" rozumiany jest jako system umieszczony w określonym środowisku, zdolny do płynnego autonomicznego działania (responsywnego i proaktywnego) w tym środowisku oraz do interakcji z ludźmi lub innymi agentami w celu osiągania zaprojektowanych zadań. Taki system może działać bez bezpośredniej interwencji człowieka i ma kontrolę nad własnym stanem i działaniami; por. *ibidem*.

[175] A. Krasuski, *op. cit.*

[176] *Ibidem*.

[177] SI cechuje dodatkowo zdolność do podejmowania samodzielnych decyzji; por. P.M. Wiórek, *Pojęcie sztucznej inteligencji*, [w:] *Sto lat polskiego prawa handlowego. Księga jubileuszowa dedykowana Profesorowi Andrzejowi Kidybie*, red. M. Dumkiewicz, J. Szczotka, K. Kopaczyńska-Pieczniak, t. 1, Warszawa 2020, LEX/el.

[178] M. Hildebrandt, K. De Vries, *Privacy, Due Process and the Computational Turn The Philosophy of Law Meets the Philosophy of Technology*, London 2015.

[179] J. Popper *et al.*, *Artificial Intelligence across Industries*, IEC White Paper 2018, s. 35.

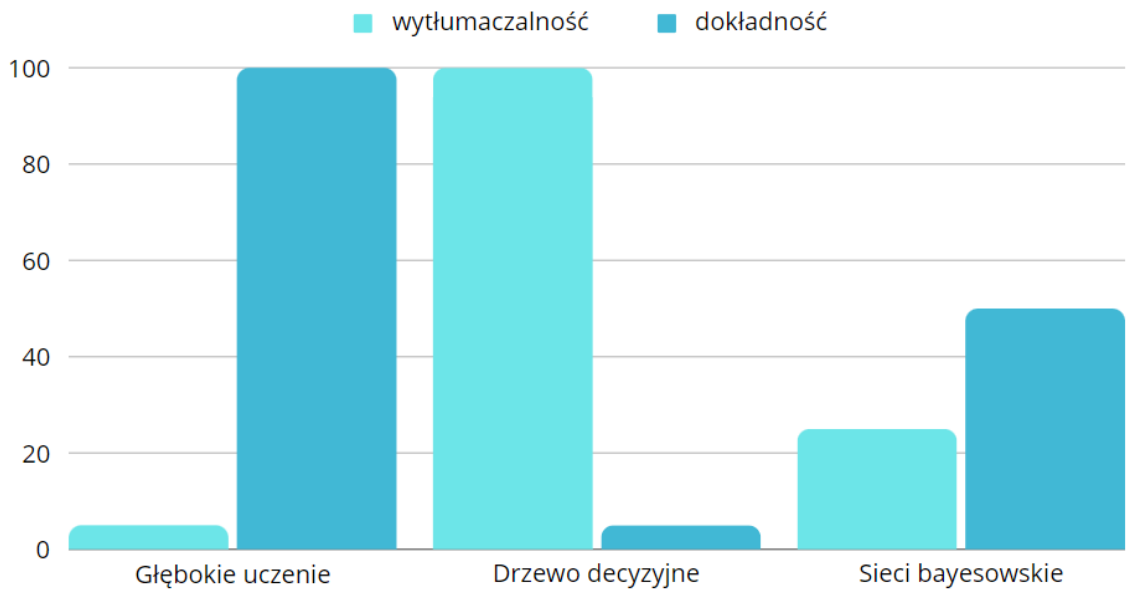
[180] P. Skalfist, D. Mikelsten, V. Teigens, *Sztuczna inteligencja: czwarta rewolucja przemysłowa*, przeł. C.S.B. Equipment, [b.m.w.] 2020, s. 33–34.

57:60414

nieustrukturyzowanych danych w nieprzetworzonej formie (np. zawartych w tekście lub obrazach), a w jego toku może dochodzić do automatycznego określania zestawów cech, które rozróżniają kategorie danych[181].

Trzecie pojęcie – sieci neuronowe – związane jest z głębokim uczeniem maszynowym. Systemy do tego rodzaju uczenia mogą bowiem wykorzystywać sieci, czyli struktury matematyczne analizujące zbiory danych, pod kątem wszystkich możliwych, optymalnych rozwiązań. Sieć neuronowa, która składa się z więcej niż trzech warstw – włącznie z warstwą wejściową i wyjściową – może być uznana za algorytm głębokiego uczenia lub głęboką sieć neuronową[182]. Struktury sieci neuronowych mają przy tym odwzorowywać zasady funkcjonowania ludzkiego mózgu.

Z procesem uczenia maszynowego związany jest jednak problem udoskonalania się (ciągłej adaptacji do danych) algorytmu w trakcie jego stosowania, co utrudnia, a nawet uniemożliwia antycypowanie efektów stosowania omawianych systemów na etapie ich tworzenia[183].



Rys. 1. Wytłumaczalność a metody uczenia – model uproszczony

[181] L. Fridman, *Deep Learning Basics: Introduction and Overview*, Massachusetts Institute of Technology, Cambridge, MA 2019, https://www.slideshare.net/noumfone/mit-deep-learning-basics-introduction-and-overview-by-lex-fridman (dostęp: 4.12.2021).
[182] *Ibidem*.
[183] J. Burrell, *How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms*, „Big Data & Society" 3, 2016, nr 1, s. 1–12.

58:47272

Źródło: opracowanie własne na podstawie S. Shirrell, *The Privacy Pro's Guide to Explainability in Machine Learning*, IAPP, https://iapp.org/news/a/the-privacy-pros-guide-to-explainability-in-machine-learning/ (dostęp: 4.12.2021).

Z dotychczasowych rozważań o charakterze pojęciowym wynika, że analizowane tu systemy dzięki zdolności uczenia się i co do zasady ograniczonej swobodzie w podejmowaniu decyzji będą mogły samodzielnie i niezależnie od wyobrażenia twórców systemu rozwiązywać problemy określonego rodzaju[184]. Z tej przyczyny sama moralność twórców, a nawet regulacje prawne, jakim podlegają, mogą nie być najważniejszym narzędziem regulacyjnym w kontekście decyzji algorytmicznych i ewentualnych błędów statystycznych. Logika regulowania dostawców ASD oraz ich operatorów, mimo że stanowi domyślny sposób wpływania na zachowania uczestników obrotu, w zderzeniu z autonomicznym generowaniem rozstrzygnięć na podstawie analizy danych populacyjnych, mogących utrwalać historyczne procesy marginalizacji i wykluczenia, dostrzega tylko fragment problemu, pomijając jego istotę.

Wniosek o nowości problemu dodatkowo uzasadnia funkcjonowanie algorytmów na podstawie ich podziału na systemy białych albo czarnych skrzynek (ang. *white box*/*black box*). W przypadku systemów zaliczanych do białych skrzynek wyniki analizy algorytmicznej są zrozumiałe dla ekspertów w danej dziedzinie, podczas gdy rezultaty procesów opierających się na systemach czarnoskrzynkowych na obecnym etapie postrzega się jako niewytłumaczalne. Problem z zagadkowością działania czarnej skrzynki wynika przy tym z metodyki funkcjonowania sztucznych sieci neuronowych i głębokiego uczenia. Jako że dane wejściowe do systemu przetwarzane są pomiędzy kolejnymi warstwami (dzięki czemu SI uczy się obserwowanych wzorców), operator systemu nie widzi danych wyjściowych wynikających z analizy wzorów dokonywanej pomiędzy warstwami, a także nie zna logiki korelacji. Brak przejrzystości w algorytmach uczenia maszynowego jest zatem wynikiem wielowymiarowości danych, złożonego kodu algorytmów i zmiennej oraz niedostrzegalnej logiki podejmowania decyzji przez taki system[185].

Tym samym, niezależnie od już dostrzeżonych problemów z małą podatnością ASD na nadzór człowieka, ryzyko błędów wzrasta w przypadku zaawansowanych algorytmów sztucznej inteligencji operujących na głębokim uczeniu maszynowym i sztucznych sieciach

[184] M. Rojszczak, *Prawne aspekty systemów*…, s. 6.
[185] S. Barocas, A.D. Selbst, *Big Data's Disparate Impact*, Social Science Research Network, Rochester, NY 2016, s. 680, https://www.cs.yale.edu/homes/jf/Barocas-Taxonomy.pdf (dostęp: 4.12.2021).

59:76511

neuronowych. Delegowanie publicznoprawnych kompetencji decyzyjnych do ASD nie może być przez to bezrefleksyjne czy nadmiernie uproszczone. Eksploracja danych uzależniona jest bowiem od jakości zbiorów, często składających się z miliardów rekordów[186]. Z jednej strony *data science* opiera się na różnych metodach analitycznych, które pozwalają wykrywać i tłumaczyć istniejące zależności między konkretnymi informacjami w danym zbiorze danych (tzw. analiza deskrypcyjna), z drugiej zaś znana jest metoda pozwalająca na przewidywanie występowania pewnych zależności na podstawie istniejących przykładów (tzw. analiza predykcyjna)[187].

Skoro systemy eksplorujące dane uczą się na przykładach, to schemat, który model zaobserwuje w danych, uzna on za statystycznie najbardziej prawdopodobny (normę)[188]. Nawet jeżeli dane będą zawierać rozstrzygnięcia obarczone jakąś formą uprzedzenia, to i tak będą służyły jako relewantne odniesienie dla dalszego procesowania. Eksploracja danych z konieczności będzie zatem wnioskować o regułach, które wykazują to samo uprzedzenie co leżące u podstaw historycznie utrwalonych schematów marginalizacji i wykluczenia[189]. Z tą jednak różnicą, że w przeciwieństwie do źródłowych procesów wykluczenia algorytmy będą działały w sposób niedostrzegalny dla jednostki i ogółu społeczeństwa. Automatyzacja procesów decyzyjnych może więc przekształcić widoczne uprzedzenia prowadzące do dyskryminacji w niewidoczną sformalizowaną regułę decyzyjną. Takie ukształtowanie nowej logiki dyskryminacji pogarszałoby szanse zarówno grup wykluczonych, jak i całej zbiorowości na przerwanie spirali marginalizacji[190]. Co więcej, eksploracja danych może nie tylko odziedziczyć wcześniejsze uprzedzenia, ale też umacniać ich aktualny wymiar, a nawet tworzyć nowe ich rodzaje[191]. Uprzedzenia te mogą być wzmacniane przy kolejnych decyzjach automatycznego systemu, co przypomina mechanizm sprzężenia zwrotnego. O ile bowiem algorytm zaobserwował większe prawdopodobieństwo słusznej decyzji na gruncie jakichś historycznie obciążonych wskaźników, o tyle wyniki takiego procesu mogłyby w sposób

[186] M. Rojszczak, *Prawne aspekty systemów…*, s. 4.
[187] M. Szeliga, Data Science *i uczenie maszynowe*, Warszawa 2017, s. 12–13.
[188] S. Barocas, A.D. Selbst, *op. cit.*, s. 680.
[189] *Ibidem*, s. 682.
[190] Ostatnie badania zaczęły podkreślać ten punkt. Jonas Lerman martwi się np. o „nielosowe, systemowe pomijanie ludzi, którzy żyją na marginesie wielkich danych, czy to z powodu ubóstwa, geografii, czy stylu życia, i których życie jest mniej »zinformatyzowane« niż życie całej populacji" – *idem, Big Data and Its Exclusions*, „Stanford Law Review" 66, 2013, cyt. za: S. Barocas, A.D. Selbst, *op. cit.*, s. 684.
[191] Można za Barocasem podać przykład portalu LinkedIn, który dzięki funkcji Talent Match typuje kandydatów do pracy. Zbadano, że jeżeli algorytm LinkedIn zaobserwuje, iż pracodawcy ignorują pewne grupy osób starających się o etat (np. członków grupy chronionej), Talent Match może zmniejszyć częstotliwość proponowania takich kandydatów pracodawcom. Mechanizm rekomendacji może się więc dostosować do uprzedzonych preferencji pracodawców; por. *ibidem*, s. 682.

60:10160

zaburzony potwierdzać prawidłowość wnioskowania i prowadzić do jego utrwalania – np. algorytm mógł uznać, że większa przestępczość występuje w dzielnicach zamieszkałych przez mniejszości kulturowe, ponieważ otrzymał dane treningowe historycznie obciążone problemem uprzedzeń, następnie na skutek decyzji algorytmu poszukiwanie przestępczości ograniczyło się jedynie do takich dzielnic, co ostatecznie uzasadniało błędny wniosek o statystycznie wyższej przestępczości w tych rejonach i przyczyniało się do dalszej intensyfikacji kontroli grup marginalizowanych.

Ze względu na bezpośredni związek między prawidłowością danych wejściowych a poprawnością działania algorytmów, rozumianą jako ograniczenie ryzyka naruszenia praw i wolności człowieka, na znaczeniu nabiera pochodzenie danych oraz sposób ich pozyskiwania[192]. Jakość i reprezentatywność analizowanych zbiorów może się przy tym różnić w sposób, który koreluje z członkostwem w danej grupie społecznej (np. instytucje mogą systematycznie utrzymywać mniej precyzyjne, terminowe i kompletne zapisy dla niektórych kategorii osób)[193].

Ten problem z perspektywy nierównej reprezentacji kobiet w danych opisała Caroline Criado Perez, ukazując go w książce *Niewidzialne kobiety*[194]. Również Solon Barocas dostrzega, że nawet zbiór danych z indywidualnymi zapisami o stałej wysokiej jakości może być obciążony tendencjami statystycznymi, które nie reprezentują różnych grup w odpowiednich proporcjach (stwierdzono zarówno niedoreprezentowanie, jak i nadreprezentację pewnych grup)[195].

Problem z selektywnością nie jest przy tym jedynym zagrożeniem związanym z operacjami na danych. Oddzielną kwestią dotyczącą prawidłowości przetwarzania danych jest znaczenie kontekstu pozyskania informacji do analizy. Michael Mattioli podsumował to zagadnienie, wskazując, że dane pozbawione kontekstu mogą być także odarte z ich pierwotnego sensu[196].

ASD podlegają zatem ograniczeniu wspólnemu dla wszystkich rodzajów przetwarzania danych, mianowicie zakłada się, że wynik nigdy nie będzie bardziej poprawny niż informacja

[192] D. Lubasz, *Zasady legalności, przejrzystości i minimalizacji danych w ogólnym rozporządzeniu o ochronie danych osobowych w kontekście sztucznej inteligencji*, [w:] *Prawo sztucznej inteligencji*…, s. 173–186.
[193] S. Barocas, A.D. Selbst, *op. cit*., s. 684.
[194] C. Criado Perez, *op. cit*.
[195] Ryzyko systematycznego pomijania w zbiorach danych członków grup chronionych powoduje, że osoby z tych grup podlegają mniej dokładnym klasyfikacjom lub przewidywaniom, ponieważ wskaźniki niezbędne do osiągnięcia ścisłych ustaleń były na poziomie dokładności i zasięgu zbiorów, których takie dane nie osiągają. Również nadreprezentacja może prowadzić do nieproporcjonalnie niekorzystnych wyników dla członków grup chronionych (np. na nich może się koncentrować nadzór). Nie zawsze przy tym łatwo ustalić, dlaczego pewne grupy chronione są niedoreprezentowane lub nadreprezentowane w danych; por. S. Barocas, A.D. Selbst, *op. cit*.
[196] M. Mattioli, *Disclosing Big Data*, „Minnesota Law Review" 99, 2014, s. 547.

61:19458

na wejściu. Na tym tle sformułowano zasadę, zgodnie z którą wnioski generowane przez systemy automatyczne mogą być tylko tak wiarygodne (ale także tak neutralne i reprezentatywne) jak dane, na których je oparto[197].

Niesprawiedliwość może zaistnieć również wtedy, kiedy algorytm uczenia maszynowego prawidłowo identyfikuje cechy umożliwiające zwiększenie prawdopodobieństwa wydania decyzji statystycznie poprawnej, ale w kontekstach, w których społeczeństwo uznaje takie różnicowanie za dyskryminujące. Podatność systemów na uprzedzenia może być przy tym trudna do zaobserwowania i wyeliminowania, albowiem proces kompilacji danych szkoleniowych i proces uczenia się wzorców są oddzielne, co oznacza, że kod algorytmu i dane nie są postrzegane w momencie tworzenia jako całość.

Sztuczna inteligencja, uczenie maszynowe i duże zbiory danych sprawiają, że automatyzacji procesów decyzyjnych towarzyszy pogłębiający się brak przejrzystości. Problem wynikający z nieprzejrzystości głębokiego uczenia doprowadził nawet do poszukiwania alternatyw dla rozwiązań czarnoskrzynkowych. Aktualną odpowiedzią na te potrzeby może się okazać wyjaśnialna SI (ang. *Explainable AI* – XAI, *Interpretable AI* lub *Explainable Machine Learning* – XML). W rzeczywistości jest to zestaw procedur, które mają umożliwić budowanie zaufania do systemów SI. Zgodnie z założeniami Departamentu Handlu USA wyjaśnialna SI musi spełniać cztery wymogi:

1) powinna dostarczać „dowodów lub uzasadnienia dla każdego wyniku";
2) wyjaśnienia powinny być zrozumiałe;
3) wyjaśnienie powinno zawierać opis procesu oraz wykorzystania danych do uzyskania rezultatów wyjściowych;
4) SI powinna działać tylko w warunkach, dla których została zaprojektowana, i nie generować innych „dodatkowych" wyników[198].

Wobec tego, że tak skonstruowana wyjaśnialna SI nie daje możliwości porównywalnych z głębokim uczeniem i sztucznymi sieciami neuronowymi, naukowcy równocześnie pracują nad metodami zwiększenia przejrzystości czarnych skrzynek oraz rozwijają inne modele analizy danych.

[197] B. Mittelstadt *et. al., op. cit.*
[198] P.J. Phillips *et al., Four Principles of Explainable Artificial Intelligence*, National Institute of Standards and Technology, U.S. Department of Commerce 2020, s. 2, https://nvlpubs.nist.gov/nistpubs/ir/2021/NIST.IR.8312.pdf (dostęp: 4.12.2021).

62:29605

Musimy przy tym pamiętać, że zaawansowane automatyczne systemy decyzyjne funkcjonują zgodnie z regułami, które są tak złożone, iż wymykają się kontroli zarówno samego decydenta, jak i organów odwoławczych i nie godzą się z aktualnymi wyobrażeniami ludzkimi o tym, jak wykonywana jest władza różnicowania[199]. Mechanizm działania ASD szczególnie poprzez zestawienie uczenia maszynowego z analizą dużych zbiorów danych stwarza dodatkowo możliwość pogłębiania się procesów wykluczenia grup historycznie marginalizowanych. Rozstrzygnięcia opierające się na statystyce i predykcji powielają istniejące uprzedzenia, jednak różnicowanie następuje w zadziwiająco nowej formie. Procesy te zaś – w związku ze swoją niejawnością, nieprzejrzystością i niewidocznością – umykają adresatom, a nawet autorom decyzji. Ukazana właśnie analiza prowadzi zatem do wniosku, że uwarunkowania technologiczne związane z automatyzacją w sposób kategoryczny wpływają na problem dyskryminacji, przez co relacje między rzeczywistą dominacją a wykształconą na ich tle ochroną zostają zaburzone.

3. Różnicowanie przez automatyczne systemy decyzyjne a problem automatycznej dyskryminacji

Przekonanie, że ASD są narzędziem neutralnym, i to takim, które można uznać za jednoznacznie korzystne, podlega krytyce w kręgach akademickich, przy czym lobby technologiczne nadal mocno wpływa na decydentów i ogół społeczeństwa[200]. Z tego względu niniejszy podrozdział ma na celu przegląd badań i ich wyników w kontekście zagrożeń stosowania automatycznych systemów decyzyjnych do realizacji postulatów równościowych. Jednocześnie analizę typologii ryzyk przeprowadzam w odniesieniu do warstwy pojęciowej procesów dyskryminacji, marginalizacji i wykluczenia.

Sukces ASD, który stoi za upowszechnieniem automatyzacji, polega na racjonalizacji procesów decyzyjnych przez ograniczanie statystycznego błędu decyzyjnego przy jednoczesnym zwiększeniu szybkości procesu oraz ograniczeniu potrzebnych zasobów osobowych i finansowych. Jak wcześniej wyjaśniałam, do procesów decyzyjnych wykorzystywane są modele statystyczne, które przy pomocy procedur lub instrukcji wyciągają

[199] E. Płonowska-Ziarek, *Rządy ludzkie czy algorytmiczne? O automatyzacji władzy sądzenia*, wykład mistrzowski, Uniwersytet Jagielloński, Kraków 2021.
[200] R. Abebe *et al., op. cit.*, s. 256.

63:90396

z danych źródłowych dotyczących osób czy zjawisk pożądane wzorce (korelacje)[201]. Dane będące podstawą rozstrzygnięć ASD generowane są wstępnie bezpośrednio przez ludzi lub pośrednio o ludziach z wykorzystaniem urządzeń elektronicznych oraz zestawiane są w zbiory określane przez ludzi. W danych przetwarzanych w systemach algorytmicznych zapisane są jednak uprzedzenia wobec grup marginalizowanych i wykluczanych, a także historia dyskryminacji – oba mające źródła w interakcjach społecznych, kulturze i stosunkach władzy istniejących mimo wykształcenia się i istnienia prawa antydyskryminacyjnego. Wpływ tych systemów na prawa i wolności człowieka na omawiane systemy jest zatem oczywisty.

Badania nad skutkami stosowania automatycznych systemów decyzyjnych odbywają się aktualnie w wielu obszarach. Powszechność problemu krzywdzącego różnicowania przez algorytmy zaprezentowała m.in. Carsten Orwat w raporcie sporządzonym dzięki grantowi z Niemieckiej Federalnej Agencji ds. Przeciwdziałania Dyskryminacji[202]. Odpowiednie przypadki zostały zebrane w ramach przeglądu literatury, który został przeprowadzony w okresie od lutego 2018 do lipca 2019 r., a następnie był wielokrotnie aktualizowany. Rozdział 4 raportu[203] w całości poświęcono studiom nad przypadkami niepożądanego różnicowania na skutek stosowania ASD. Opisano w nim m.in.:

1) oprogramowanie kadrowe w Amazonie, które nie było stronnicze co do płci kandydatów i pozostało tendencyjne pomimo wprowadzanych korekt;
2) platformy internetowe do wyszukiwania zleceń przez freelancerów, takie jak TaskRabbit i Fiverr – okazało się, że TaskRabbit był w znacznym stopniu skorelowany z pochodzeniem etnicznym i płcią, z kolei w wyniku stosowania Fiverr osoby czarnoskóre otrzymały m.in. mniej recenzji i niższe oceny;
3) wykluczające profilowanie adresatów ofert o pracę poprzez mikrotargetowanie (m.in. w takich firmach jak Amazon, Verizon, UPS, Goldman Sachs i Facebook) w celu ukierunkowania rekrutacji na określone grupy wiekowe;
4) predykcję użytkowników Facebooka w zakresie rodzaju wyświetlanej reklamy pod kątem najwyższego prawdopodobieństwa reakcji na komunikację;

[201] T. Gillespie, *The Relevance of Algorithms*, [w:] *Media Technologies: Essays on Communication, Materiality, and Society Get Access Arrow*, red. T. Gillespie, P.J. Boczkowski, K.A. Foot, Cambridge, MA 2014, s. 167.
[202] C. Orwat, *Risks of Discrimination through the Use of Algorithms*, Federal Anti-Discrimination Agency 2019, ttps://www.antidiskriminierungsstelle.de/EN/homepage/_documents/download_diskr_risiken_verwendung_von_algorithmen.pdf?__blob=publicationFile&v=1 (dostęp: 4.12.2021).
[203] *Ibidem*, rozdz. 4.

64:52114

5) nierówne traktowanie użytkowników Airbnb ze względu na pochodzenie etniczne – wynajmujący niebędący osobami czarnoskórymi otrzymywali wyższe ceny wynajmu za porównywalne oferty w porównaniu z czarnoskórymi wynajmującymi, a azjatyccy wynajmujący uzyskiwali niższe ceny niż biali wynajmujący za porównywalne oferty;
6) różnicowanie użytkowników ze względu na rodzaj wykorzystywanej przeglądarki internetowej, systemu operacyjnego czy wyszukiwarki, a także posiadania konta użytkownika na danym portalu lub historii zakupów;
7) nierówne traktowanie w logistyce Amazona poprzez odmowę dostawy do dzielnic zamieszkiwanych w wysokim odsetku przez osoby czarnoskóre;
8) reprodukcję stereotypów w toku generowania przez wyszukiwarki linków i obrazów lub w autouzupełnieniach słów w polu wyszukiwania;
9) dyskryminację w sektorze bankowym ze względu na pochodzenie etniczne;
10) przypadek firmy Svea Ekonomi, dotyczący wielokrotnej dyskryminacji w sektorze udzielania pożyczek online;
11) tendencyjne wyniki systemu diagnostycznego zaprojektowanego w celu przewidywania ryzyka śmiertelności pacjentów z zapaleniem płuc;
12) dyskryminację ze względu na pochodzenie etniczne i uprzedzenia na tle rasowym w systemie zarządzania placówkami medycznymi w zakresie przydzielania pacjentów wymagających intensywnej opieki;
13) system oceny kierowców przez użytkowników usługi *ride sharing*, będący źródłem dyskryminacji ze względu na pochodzenie etniczne lub rasę;
14) opisywany już w tej dysertacji polski ASD do profilowania bezrobotnych;
15) niedokładność systemów rozpoznawania twarzy, co przyczyniało się do częstszych kontroli grup marginalizowanych i wykluczanych.

Znane są również przypadki austriackiego algorytmu[204], który ograniczał szanse na zatrudnienie kobiet w wieku powyżej 30 lat, kobiet mających obowiązki w zakresie opieki nad dziećmi, migrantek lub osób z niepełnosprawnościami[205].

[204] Arbeitsmarktservice (AMS) – algorytm służący do przewidywania perspektyw zatrudnienia osoby poszukującej pracy na podstawie takich czynników jak: płeć, grupa wiekowa, obywatelstwo, stan zdrowia, zawód i doświadczenie zawodowe.
[205] D. Allhutter *et al.*, *Algorithmic Profiling of Job Seekers in Austria: How Austerity Politics Are Made Effective*, „Frontiers in Big Data" 3, 2020, nr 5.

65:24964

Podczas gdy wielu badaczy koncentruje się na wadliwych ASD, istnieją również pozytywne przykłady[206], których jednak z perspektywy problemu omawianego w niniejszej pracy nie czynię przedmiotem pogłębionej analizy.

Na tym tle można uznać za uprawniony pogląd, że skuteczność oferowana przez systemy algorytmiczne może nie przystawać do idei sprawiedliwości czy równości. Ryzyko niesprawiedliwego różnicowania pozostaje przy tym na podobnym poziomie w ASD stosowanych w sektorze zarówno prywatnym, jak i publicznym oraz dotyka przede wszystkim grup uprzednio już wykluczonych politycznie, ekonomicznie i społecznie. Jaka jest jednak przyczyna tego swoistego błędnego koła wykluczenia? Krytycy eksploracji danych koncentrują się na nierzetelnych klasyfikacjach, tzn. fałszywie pozytywnych lub negatywnych, aczkolwiek podobnie dużo niebezpieczeństwa kryje się w samej definicji klasy i późniejszym etykietowaniu przykładów, z których wnioskuje się o regułach[207]. Prawa podstawowe mogą być jednak naruszone czy to na etapie określenia kategorii bądź wyznaczania celów rozróżnienia, czy to na podstawie błędnych danych, na bazie których następuje klasyfikowanie.

Przywołując przykład opisywany przez Bettinę Berendt[208], można wyjaśnić, że algorytm, który wygeneruje wynik w postaci ustalenia, że kobiety są niższe od mężczyzn, nie jest tendencyjny, lecz statystycznie poprawny, ponieważ uwypukla biologiczną prawidłowość. Również algorytm, który sklasyfikuje kobiety jako zarabiające mniej niż mężczyźni, będzie statystycznie poprawny, więc nadal nietendencyjny. Replikowanie odzwierciedlonych skutków historycznej marginalizacji kobiet ze względu na płeć nie będzie jednak pożądane, choć właśnie taki będzie skutek, gdy wyniki analizy zostaną użyte do predykcji lub celów normatywnych[209]. Tym samym chociaż korelacje uwidoczniane na skutek przetwarzania danych w ASD mogą być statystycznie prawdziwe, to w wymiarze społecznym i politycznym uznawanie ich za opisujące normę populacyjną do budowania modeli decyzyjnych będzie szkodliwe i niesprawiedliwe. Na skutek tego sprzężenia zwrotnego ekspansja algorytmicznych narzędzi decyzyjnych może w sposób niewidoczny dla społeczeństwa nadać nową prędkość

[206] Alex P. Miller badał ASD stosowany w hrabstwie Allegheny. System ten, jako pierwszy w Stanach Zjednoczonych, poddawał decyzje pracowników Urzędu ds. Dzieci, Młodzieży i Rodzin (Children, Youth, and Families Department – CYF) dotyczące identyfikowania rodzin najbardziej potrzebujących interwencji kontroli algorytmu o nazwie Allegheny Family Screening Tool; zob. *idem*, *Decision Making and Problem Solving Want Less-Biased Decisions? Use Algorithms*, „Harvard Business Review" 2018, https://hbr.org/2018/07/want-less-biased-decisions-use-algorithms (dostęp: 4.12.2021).
[207] S. Barocas, A.D. Selbst, *op. cit*., s. 680.
[208] B. Berendt, *Algorithmic Discrimination*, [w:] *Encyclopedia of Law and Data Science*, 2020, s. 5, https://people.cs.kuleuven.be/~bettina.berendt/Papers/Algorithmic%20Discrimination%20BB%202020%2010%2013.pdf (dostęp: 4.12.2021).
[209] *Ibidem*.

66:81950

procesom wykluczenia grup marginalizowanych, a przez to pogłębić procesy dyskryminacji w niespotykanych dotąd formach.

Klasyfikowanie ludzi przy pomocy automatycznych systemów decyzyjnych może mieć istotny wpływ też na inne prawa, takie jak prawo do samostanowienia, autonomii informacyjnej czy wolność wypowiedzi i prywatności, to zaś nadaje procesom dyskryminacji nowy wymiar i format[210]. Ta nowość manifestuje się już w trudnościach z oceną prawną tego zjawiska. Omówione w pierwszym podrozdziale przypadki polskiego systemu do profilowania bezrobotnych oraz niderlandzkiego systemu SyRI naruszały liczne gwarancje państwa demokratycznego, jednak decyzje sądów wstrzymujące stosowanie tych systemów nie opierały się w relewantnym zakresie na zakazie dyskryminacji przez te ASD. Tymczasem oba wskazane przykłady pogłębiały istniejące procesy marginalizacji i wpływały na wykluczenie określonych grup społecznych.

Już na tej podstawie można dojść do wniosku, że nie każda forma oceny i kategoryzacji algorytmicznej będzie spełniała ustawowe przesłanki dyskryminacji, chociaż jednocześnie będzie prowadziła do wykluczenia i marginalizacji. Dyskryminacja jest bowiem kwalifikowaną formą różnicowania i nie polega wyłącznie na klasyfikacji, nawet gdy kategoryzowanie jest błędne. W przypadku ASD problem dyskryminacji jest jednak szerszy i wykracza poza samą ocenę różnicowania z perspektywy realizacji postulatów równościowych. Również mechanizm i narzędzia wykorzystywane do procesu decyzyjnego są inne, przez co procesy marginalizacji prowadzące do dyskryminacji są ukryte i niezrozumiałe dla ludzi, a przy tym masowe na skalę globalną.

Wobec tego dalszą analizę sugeruję poprzedzić ustaleniem relacji między pojęciami dyskryminacji i wykluczenia, w tym z perspektywy algorytmicznej (ang. *algorithmic bias*). Określenie *bias* jest obecne od dawna[211] w dziedzinie uczenia maszynowego i zwyczajowo odnosi się do założeń przyjętych przez dany model (ang. *inductive bias*)[212]. Angielski termin *bias* nie ma niestety jednolitego tłumaczenia na język polski, przez co może być rozumiany jako uprzedzenie, wykluczenie, jak również stronniczość czy skrzywienie. Stronniczość ma przy tym szersze znaczenie niż dyskryminacja, ponieważ dotyczy wszystkich rodzajów systematycznych błędów, które mogą obejmować te o charakterze statystycznym,

[210] O. Pollicino, G. De Gregorio, *Constitutional Law in the Algorithmic Society*, [w:] *Constitutional Challenges in the Algorithmic Society*, red. H.W. Micklitz *et al.*, Cambridge, MA 2021, s. 3–24.
[211] W naukach społecznych Helen Nissenbaum pisała o stronniczej technologii już w drugiej połowie lat 90. XX w.; zob. *Computers, Ethics & Social Values*, red. D.G. Johnson, H. Nissenbaum, Upper Saddle River, NJ 1995.
[212] T.M. Mitchell, *Machine Learning*, New York 1997.

67:54497

poznawczym, społecznym, strukturalnym lub instytucjonalnym, na skutek których grupy nieuprzywilejowane w systematycznej przewadze traktowane są niekorzystnie w stosunku do grup uprzywilejowanych[213]. Określeniem *bias* opisujemy zaś utrwalone błędy obserwowane w algorytmie, występujące samodzielnie lub jako element automatycznego systemu decyzyjnego. Błędy mogą przy tym powodować niesprawiedliwe wyniki, takie jak uprzywilejowanie jednej kategorii względem drugiej w sposób inny niż zamierzona funkcja algorytmu.

Rozpoznawane są różne powody wykluczenia algorytmicznego[214]. W raporcie Biura Wykonawczego Prezydenta Stanów Zjednoczonych *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights*[215] wskazano najistotniejsze przyczyny obserwowanych nieprawidłowości:

1) dane mogą być źle dobrane (dane nieistotne będą miały wpływ na wynik procesu);
2) dane mogą być niekompletne, nieprawidłowe lub nieaktualne;
3) zestaw danych wejściowych do modelu może być niereprezentatywny dla populacji, a zatem wyniki również będą tendencyjne (będą faworyzowały pewne grupy);
4) dane będą utrwalać w sposób niezamierzony lub celowy uprzedzenia historyczne[216].

Taka stronniczość ASD może mieć charakter zarówno krzyżowy, jak i emergentny. Metody ją wspierające będą jednak powszechnie zyskiwały walor cyfrowej neutralności i gospodarczej efektywności. W rzeczywistości zaś procesy rozkładu społecznego będą postępowały masowo, a niewidzialne, a nawet niezrozumiałe dla człowieka różnicowanie nie będzie przedmiotem społecznej odpowiedzi. Efektem tego zjawiska będzie – podobnie jak tych opisanych w rozdziale I – uprzywilejowanie jednej grupy ludzi względem innych. Taka dominacja może mieć przy tym charakter nam znany, polegający na pogarszaniu sytuacji grup społecznych

[213] J. Gerards, R. Xenidis, *Algorithmic Discrimination in Europe: Challenges and Opportunities for Gender Equality and Non-Discrimination Law*, European Commission, Brussels 2020, https://op.europa.eu/en/publication-detail/-/publication/082f1dbc-821d-11eb-9ac9-01aa75ed71a1 (dostęp: 4.12.2021).
[214] Są to np. stronniczość próby (ang. *sample bias*), stronniczość wykluczenia (ang. *exclusion bias*), stronniczość pomiaru (ang. *measurement bias*), stronniczość przywołania (ang. *recall bias*), stronniczość obserwatora (ang. *observer bias*), stronniczość rasowa (ang. *racial bias*) czy stronniczość powiązania (ang. *association bias*).
[215] C. Muñoz, M. Smith, D. Patil, *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights*, Executive Office of the President 2016, https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf (dostęp: 4.12.2021).
[216] *Ibidem.*

68:10652

historycznie marginalizowanych, ale również nowy – tworzący inne ośrodki władzy i nowe grupy algorytmiczne[217].

Na ogół marginalizacja wynikała z braku dostatecznej reprezentacji danych grup w organach politycznych stanowiących prawo. Niedysponowanie reprezentacją czyniło zaś te grupy niewidocznymi na niesprawiedliwość wynikającą z egzekwowania przepisów w takich strukturach, przez co wykluczenie stawało się zarówno przyczyną, jak i skutkiem konkretnych procesów. W przypadku algorytmów marginalizacja również będzie wynikać z interesów gospodarczych. Grupy uciskane mogą bowiem pozostawać poza istotnym zainteresowaniem lobby technologicznego, przez co nie będą brane pod uwagę przy ustalaniu wzorców oraz zabezpieczaniu ich praw i wolności. Opisane procesy nie są przy tym równoznaczne z różnicowaniem zarówno bezpośrednim, jak i pośrednim na podstawie cech chronionych prawem antydyskryminacyjnym. Co więcej, kategoryzowanie algorytmiczne nie jest widoczne dla ludzi i przez to trudniej o społeczny czy polityczny opór wobec pogłębiających się klasyfikacji i nierówności.

Można przywołać w tym kontekście opisywane systemy SyRI i polski ASD do profilowania bezrobotnych. Z uwagi na nieprzejrzystość obu systemów wyniki rozumiane jako jednostkowe decyzje ASD nie pozwalały dostrzec szerszej perspektywy oddziaływania automatyzacji procesu decyzyjnego na społeczeństwo. Dopiero całościowa analiza szeregu decyzji wydawanych przez ASD na skutek działań organizacji pozarządowych umożliwiła stwierdzenie dużego prawdopodobieństwa systematycznego pogarszania sytuacji jednostek z grup historycznie marginalizowanych. Ryzyko wykluczenia nie zostało dostrzeżone przez organy państwowe i nie było badane na etapie tworzenia przepisów, a na skutek braku przejrzystości i jawności procesów różnicowania reakcja adresatów jednostkowych decyzji ASD, społeczeństwa i organów kontroli była spóźniona.

Uwzględniając dotychczasowe rozważania, należy stwierdzić, że wykluczenie algorytmiczne jest pojęciem szerszym niż dyskryminacja. Błąd algorytmiczny może, jak w przypadku SyRI lub polskiego ASD do profilowania bezrobotnych, nie tylko wynikać z kategoryzacji na podstawie zestawu danych zgromadzonych w systemach i bazach danych państwowych, ale też z klasyfikowania ludzi poprzez analizę innych ich atrybutów, nieutożsamianych z elementami chronionymi w ramach godności człowieka (takich jak dochód, kod pocztowy,

[217] S. Wachter, *The Theory of Artificial Immutability: Protecting Algorithmic Groups under Anti-Discrimination Law*, „Tulane Law Review" 20, 2022, nr 97, s. 1–50.

69:37114

historia przeglądarki, zapisy z urządzeń IoT itp.) lub ze złożonych algorytmicznych kombinacji kilku atrybutów i/lub cech ludzkich[218]. Warto podnieść, że profile skonstruowane przez algorytmy na podstawie neutralnych cech czy atrybutów, takich jak kod pocztowy, mogą (nawet niecelowo) pokrywać się z innymi profilami związanymi z pochodzeniem etnicznym, płcią, orientacją seksualną itd.[219] Mittelstadt słusznie zauważa, że od samego początku pozostaje niejasne, jakie rodzaje łączników tożsamości behawioralnej oraz modeli decyzyjnych można wytworzyć i które z nich są potencjalnie kłopotliwe etycznie, a w konsekwencji, które obszary stosowania ASD powinny podlegać ograniczeniom lub regulacji[220].

Okoliczność ta jest o tyle istotna, że zaawansowane ASD będą prowadziły wnioskowanie i wydawały rozstrzygnięcia bez relewantnego udziału ludzkiego decydenta i poza jego kontrolą i rzeczywistym nadzorem. Analityka danych będzie zastępować świadome projektowanie reguł kontroli procesów decyzyjnych, wypierając metody badania empirycznego, przechodząc od przyczynowości jako powiązania między danymi wejściowymi i wyjściowymi do korelacji[221].

Ten konflikt między antydyskryminacyjnymi ramami, które pojmują wiedzę jako ustalanie przyczynowości, a analityką opartą na płynnych hipotezach świadczy o istotnej zmianie okoliczności stosowania prawa[222]. Nieprzejrzystość epistemiczna systemów, niejawność procesów różnicowania, niewyjaśnialność czarnych skrzynek, błędy w zbiorach danych, a także ich selektywność są prawdziwym problemem, mogącym mieć zasadniczy wpływ na realizację zakazu dyskryminacji. Znaczenie danych skonstruowane poprzez ich obserwacje niekoniecznie będzie prawidłowe czy wytłumaczalne, ale może niezaprzeczalnie racjonalizować to, co w przeciwnym razie byłoby uznane za przypadkowe i niegodne powielania[223]. Pozostawienie technologii, w tym automatycznych systemów decyzyjnych, poza ramami demokratycznych systemów wartości może ponadto prowadzić społeczeństwo do technodeterminizmu i marginalizacji władzy publicznej, która utraciłaby swoją rolę w realizacji ochrony praw podstawowych[224]. Specyfika działania algorytmów może przy tym

[218] R. Gellert *et al.*, *A Comparative Analysis of Anti-Discrimination and Data Protection Legislations*, [w:] *Discrimination and Privacy in the Information Society. Studies in Applied Philosophy, Epistemology and Rational Ethics*, red. B. Custers *et al.*, t. 3, Berlin-Heidelberg, s. 80.
[219] B.D. Mittelstadt, *From Individual to Group Privacy in Big Data Analytics*, „Philosophy & Technology" 30, 2017, s. 479.
[220] *Ibidem*.
[221] M. Mattioli, *op. cit.*, s. 538.
[222] M. Leese, *The New Profiling: Algorithms, Black Boxes, and the Failure of Anti-Discriminatory Safeguards in the European Union*, „Security Dialogue" 45, 2014, s. 505.
[223] A. Jacobs, *The Pathologies of Big Data*, „Queue" 7, 2009, nr 6, s. 10–19.
[224] O. Pollicino, G. De Gregorio, *op. cit.*, s. 5.

70:10351

umacniać, a nawet zmieniać relacje władzy przez wykorzystanie baz danych tworzonych o obywatelach do śledzenia i nadzoru[225]. Automatyczne narzędzia zarządzania algorytmicznego mogą być bowiem wprowadzane na masową skalę pod pretekstem ulepszenia administracji publicznej, promując jej bardziej efektywne, dokładne i sprawiedliwe formy działania. Takie stosowanie metod statystycznych wzmaga jednocześnie nadzór predyktywny, uprzedzający i prewencyjny[226]. Wobec niewidzialności procesów różnicowania prowadzących do wykluczenia pewnych grup wraz z tą ekspansją wzrasta ryzyko eksplorowania danych w celu obejścia barier, zarówno praktycznych, jak i prawnych, poprzez ukrycie różnicowania opartego pośrednio na cesze chronionej prawem antydyskryminacyjnym. Co za tym idzie – brak faktycznego ludzkiego nadzoru nad procesem różnicowania, jego epistemiczna nieprzejrzystość i technologiczna niewyjaśnialność wpływają na niepowodzenia tradycyjnych form reakcji społecznej na niesprawiedliwość związaną z wykluczeniem i marginalizacją.

Błędy w systemach automatycznych mogą przy tym na niespotykaną skalę potęgować istniejące nierówności społeczne, a nawet tworzyć nowe jej oblicza. Analiza dużych zbiorów danych może bowiem istotnie zwiększyć i ułatwić ukrycie aktów zamierzonej dyskryminacji poprzez posługiwanie się kryteriami bardziej odległymi od cech chronionych, a także ich substratami[227]. Co więcej, jak wskazuje Cathy O'Neil[228], w miarę postępu technologicznego będziemy sortowani w nowe rodzaje grup, które jednak nie będą oparte na tradycyjnych cechach, takich jak płeć czy wiek, a na danych behawioralnych[229]. Doniosłość problemu badawczego jest więc znaczna zarówno z perspektywy społecznej, jak i prawnej.

Opisana już trudność w zaobserwowaniu procesu niesprawiedliwego algorytmicznego różnicowania w trakcie jego trwania doprowadziła do tego, że wykrycie błędów algorytmicznych w większości przypadków było związane z zaangażowaniem instytucji

[225] E. Ntoutsi *et al.*, *op. cit.*, s. 4.
[226] D. Murakami Wood, *op. cit.*
[227] S. Barocas, A.D. Selbst, *op. cit.*, s. 693.
[228] C. O'Neil, *op. cit.*
[229] Warto przywołać tu opisywany przez Cathy O'Neil przypadek systemu decyzyjnego badanego przez ekspertów z Sense. Organizacja ta rozpoczęła analizowanie danych z telefonów komórkowych, pokazujących, jak ludzie się przemieszczają. Dane do analizy pochodziły z USA i Europy; informacje zostały zanonimizowane, przez co analizie podlegały kropki przemieszczające się po mapie. Zatrudniony w tej firmie zespół analityków wprowadził do systemu uczenia maszynowego dane bez dodatkowych wskazówek. Program miał samodzielnie dokonać sortowania danych na kategorie, które sam opracował. Badacze nie polecili mu przy tym, by oddzielał mieszkańców dzielnic podmiejskich czy nastolatków ani by tworzył odrębne grupy o podobnych schematach zakupowych. Oprogramowanie miało działać autonomicznie. Na skutek analizy punkty na mapie zaczęły przybierać różne kolory – jedne stawały się czerwone, natomiast inne żółte, niebieskie czy zielone. Grupy kolorów nie były jednak zrozumiałe dla człowieka. Współzałożyciel Sense stwierdził, że osoby te nie pasowały do żadnych tradycyjnych kategorii. Więcej zob. *ibidem*.

71:94811

międzynarodowych bądź organizacji społecznych w proces ustalania przyczyn obserwowanych nieprawidłowości. Dla indywidualnego adresata decyzji algorytmicznych, jak również dla operatora systemu, który był zaangażowany w pojedynczy proces decyzyjny, naruszenia nie były bowiem możliwe do zidentyfikowania. Równocześnie twórcy systemów i korzystające z nich podmioty rzadko zlecały analizy efektów ubocznych procesu decyzyjnego.

Jak pisała C. O'Neil w książce *Weapons of Math Destruction*[230], obecnie stosowane modele matematyczne są niezrozumiałe, a sposób ich działania są w stanie zgłębić jedynie „najwyżsi kapłani – matematycy i programiści komputerowi". Systemy decyzyjne zawierają zatem opinie ludzkie napisane językiem matematyki. W ASD funkcjonują nieracjonalne uprzedzenia, opierające się na historycznych uwarunkowaniach społecznych. Rozstrzygnięcia ferowane przez te modele, nawet gdy niosą z sobą społeczną szkodliwość, nie podlegają otwartej dyskusji ani krytyce z powodu ich domniemanej technologicznej bezstronności, a jednocześnie utajnienia procesów różnicowania. Same zaś korelacje, jak wskazywała autorka, mają tendencję do pogarszania losu osób ubogich i wykluczonych. Odmiennie do przykładów historii walki o równość ofiary nie wiedzą w tym wypadku, że ich prawa i obowiązki kreowane są pod wpływem mikrotargetowania[231] czy skomplikowanych technik analizy dużych zbiorów danych. Natomiast brak przejrzystości towarzyszący systemom automatycznego podejmowania decyzji prowadzi do tego, że nawet świadomość błędności decyzji czy jej niesprawiedliwości (np. nieuzyskanie kredytu, miejsca na uniwersytecie czy stanowiska pracy) nie umożliwia skutecznego poznania rzeczywistej przyczyny nieprawidłowości, a w konsekwencji nie daje argumentów do merytorycznego zakwestionowania decyzji[232]. Adaptacyjna i wyprzedzająca percepcję człowieka predyspozycja ASD może w związku z tym pozbawić jednostki realnej możliwości, a nawet umiejętności rozumowania zgodnie z nowymi regułami decyzyjnymi i kryteriami[233], a przez to uniemożliwić właściwą i szybką reakcję zarówno obywateli, jak i władzy publicznej na rozkład społeczny spowodowany trwającymi procesami wykluczenia algorytmicznego.

Takie *novum* w procesach różnicowania wymaga również refleksji nad adekwatnością narzędzi wykształconych w ramach prawa antydyskryminacyjnego. To zaś w konsekwencji może

[230] *Ibidem* (wyd. pol. C. O'Neil, *Broń matematycznej zagłady: jak algorytmy zwiększają nierówności i zagrażają demokracji*, przeł. M.Z. Zieliński, Warszawa 2020). Warto nadmienić, że autorka jest matematyką, badaczką danych i współzałożycielką portalu www.mathbabe.org.
[231] Mikrotargetowanie oznacza dostosowywanie wyświetlanych treści do indywidualnych profili użytkowników.
[232] J. Niklas, *Problem dyskryminacji automatycznej – uwagi na tle ogólnego rozporządzenia o ochronie danych osobowych*, „Europejski Przegląd Sądowy" (dalej: „EPS") 7, 2019, s. 4–9.
[233] E. Bayamlıoğlu, R. Leenes, *op. cit.*

72:19642

prowadzić do konieczności zrewidowania przekonania, że sprawiedliwość można osiągnąć dzięki prostej zmianie sposobu oceniania ludzi w poszczególnych momentach procesu decyzyjnego[234].

Analizowane przeze mnie procesy nie były wolne od dyskryminacji, wykluczenia i marginalizacji przed pojawieniem się i upowszechnieniem automatycznych systemów decyzyjnych, jednak charakter tych zagrożeń był inny. Tymczasem cyfryzacja kształtuje sposób, w jaki problemy społeczne są redefiniowane, zmieniając rozumienie tych problemów i możliwych odpowiedzi na nie[235]. Musimy też brać pod uwagę, że automatyzacja to także narzędzie do wyświetlenia fałszywych korelacji, dzięki czemu możliwe będzie lepsze przeciwdziałanie wykluczeniu i marginalizacji na wszelkich polach społecznych. Jedną z głównych przyczyn dyskryminacji jest bowiem błędna wiara w statystyczne zależności, które nie mają podstaw w rzeczywistości[236]. Badania nad ASD zwiększyły tymczasem nasz potencjał odtajniania i dokumentowania wzorców dyskryminacyjnych w poszczególnych systemach socjotechnicznych[237]. Wiedzę tę należy wykorzystać w celu lepszego opracowania metody walki z irracjonalnymi uprzedzeniami.

Jak wskazał statystyk George Box: „Wszystkie modele są błędne, ale niektóre są użyteczne”[238]. O ile zatem zgodnie ze słowami Boxa dostrzeżemy błędy algorytmów, o tyle będziemy mogli się na tych błędach nauczyć przeciwdziałać dyskryminacji na innym poziomie. To zaś może się okazać punktem zwrotnym w walce o realizację postulatów równościowych.

[234] S. Barocas, M. Hardt, A. Narayanan, *Fairness and Machine Learning*, https://fairmlbook.org (dostęp: 4.12.2021).
[235] R. Abebe *et al.*, *op. cit.*, s. 252.
[236] F.K. Thomsen, T. Khaitan, F. Schauer, *op. cit.*, s. 47.
[237] Cyt. za: R. Abebe *et al.*, *op. cit.*, s. 253.
[238] G. Box, *Robustness in the strategy of scientific model building*, Robustness in Statistics, red. Robert Launer i Graham Wilkinson, Academic Press, 1979, s. 202.

73:68104

III. Europejskie regulacje prawne jako źródło zakazu różnicowania w algorytmicznych procesach decyzyjnych

Technologie cyfrowe prowadzące do automatyzacji procesów decyzyjnych, wykorzystywane przez nie dane oraz narzędzia do ich przetwarzania, a także efekty ich stosowania muszą odpowiadać wartościom i zasadom leżącym u podstaw umowy społecznej oraz reguł prawnych, które ją formalizują. Zaprezentowana w tym rozdziale analiza przede wszystkim odwołuje się do regulacji europejskiego prawa antydyskryminacyjnego, jako że sprawiedliwość algorytmów badana jest głównie z perspektywy założeń dozwolonego różnicowania. Treść zakazu dyskryminacji przyjęta w Europejskiej konwencji o ochronie praw człowieka i podstawowych wolności oraz w ramach późniejszych aktów Unii Europejskiej, zwłaszcza Karty Praw Podstawowych, będącej jednym z najważniejszych narzędzi ochrony praw podstawowych, wyznacza przy tym granice owego zakazu w regulacjach szczególnych, takich jak ochrona konsumentów, pracowników czy dostęp do edukacji. Następnie, o ile niesprawiedliwe różnicowanie specyficzne dla algorytmów nie będzie w pełni objęte zakazem dyskryminacji, należy zbadać inne szczegółowe regulacje dotyczące automatycznego przetwarzania danych, w tym w ramach profilowania znanego z ochrony danych osobowych lub prawa do autonomii informacyjnej, gwarantowanego prawem do prywatności.

Należy przy tym podkreślić, że zakaz dyskryminacji jest jednym z filarów nowożytnego państwa prawa[239], dlatego też niniejsza analiza dotyczy w głównej mierze regulacji praw człowieka jako najdalej idących w swoich gwarancjach. Z powodu wielu pojęciowych wyzwań opisanych w rozdziale I niniejszej pracy, które charakteryzowały definicję analityczną pojęcia dyskryminacji od samego początku jej wykształcenia jako „formalnego pojęcia w prawie" i w etyce, badanie specyfiki problemu algorytmicznego podejmowania decyzji przebiega na granicy prawa antydyskryminacyjnego i prawa równościowego w szerokim znaczeniu[240]. Celem tego rozdziału jest analiza możliwości zakwalifikowania różnicowania algorytmicznego jako formy dyskryminacji, a w wypadku braku takiej możliwości – poszukiwanie

[239] Żadne różnice w traktowaniu, które wynikają wyłącznie lub w decydującym stopniu z pochodzenia etnicznego danej osoby, nie mogą być obiektywnie uzasadnione we współczesnym społeczeństwie demokratycznym, opartym na zasadach pluralizmu i poszanowania odmiennych kultur; zob. wyrok ETPCz z 22.12.2009 r. w sprawie *Sejdić i Finci v. Bośnia i Hercegowina*, skargi nr 27996/06 i 34836/06, par. 44. Podobnie: wyrok ETPCz z 13.12.2005 r. w sprawie *Timishev v. Rosja*, skargi nr 55762/00 i 55974/00, par. 58.

[240] S. Barocas, *Data Mining and the Discourse on Discrimination*, Center for Information Technology Policy, Princeton University, Princeton, NJ 2014, s. 3–4, https://www.cs.yale.edu/homes/jf/Barocas-Taxonomy.pdf (dostęp: 4.12.2021).

74:10889

komplementarnych regulacji prawnych gwarantujących ochronę przed wykluczeniem i marginalizacją.

1. Historia i treść prawnego zakazu dyskryminacji

Stosunek władzy ustawodawczej do problematyki nierównego traktowania zmieniał się wraz z przemianami, które zachodziły w społeczeństwach na skutek walk i konfliktów, a także niejako równolegle w odpowiedzi na roszczenia różnych grup w kwestii poszanowania ich podstaw prawnych oraz zmieniających się zasad i wartości kształtujących demokratyczno-liberalne społeczeństwa.

Prawo antydyskryminacyjne (ang. *anti-discrimination law*, niem. *Antidiskriminierungsgesetz*, franc. *loi anti-discrimination*) to stosunkowo nowa dziedzina prawa prywatnego, która reguluje zagadnienia związane z prawem do równego traktowania i ochroną przed dyskryminacją w różnych sferach życia społecznego[241]. Jego adresatami są podmioty zarówno prywatne, jak i publiczne, przy czym historycznie zakaz miał sformalizować ograniczenia władzy w kontekście wprowadzonych lub utrwalonych przez nią stosunków dominacji. W odpowiedzi na rozmaite przejawy naruszenia równości i godności upubliczniane w toku zmian rewolucyjnych oraz ze względu na pojedyncze sprawy sądowe czy zmiany legislacyjne wynikające z przekształceń politycznych prawo często przewodziło powszechnemu rozumieniu tego, co oznacza dyskryminacja[242]. Regulacja zakazu dyskryminacji wydaje się przy tym zbiorem kilku fragmentów i reżimów prawa, nie zaś jednolicie wykształconą, autonomiczną dziedziną. Z jednej strony niektóre aspekty prawa antydyskryminacyjnego są bowiem związane z konstytucją i prawami człowieka (prawami publicznymi, politycznymi), z drugiej – historyczne korzenie prawa antydyskryminacyjnego są ściśle połączone z rozwojem prawa pracy, a następnie ochroną konsumentów i wolnego rynku. Warto też nadmienić, że dyskryminacja jako czyn może być utożsamiana z deliktem.

Taka złożoność samej dyscypliny jest dodatkowo ubarwiana wielością źródeł regulacji, jako że na system ogólnego prawa antydyskryminacyjnego składają się przepisy prawa zarówno krajowego, jak i międzynarodowego, w tym europejskiego. Z krajowego punktu widzenia

[241] J. Maliszewska-Nienartowicz, *Prawo antydyskryminacyjne*, red. J. Maliszewska-Nienartowicz, „System Prawa Unii Europejskiej", t. 6, Warszawa 2020, s. 2.
[242] F.K. Thomsen, T. Khaitan, F. Schauer, *op. cit.*, s. 30.

75:11094

najważniejszym aktem prawnym obowiązującym w tym zakresie jest Konstytucja Rzeczypospolitej Polskiej z 2.02.1997 r.[243] Niemniej jednak z perspektywy analizy zagadnienia równie istotną rolę dla obecnego rozumienia pojęcia prawnego dyskryminacji odgrywają akty prawa międzynarodowego przyjęte przez Organizację Narodów Zjednoczonych i Radę Europy oraz prawo antydyskryminacyjne rozwijane w ramach Unii Europejskiej.

Mimo że regulacje zakazujące dyskryminacji na różnych polach są dość młode, ich tworzeniu towarzyszyła rosnąca świadomość społeczna związana z koniecznością zapobiegania zjawisku nieuzasadnionej dominacji i marginalizacji. Był to wynik sprzeciwu uciskanych grup społecznych bądź szerszych zjawisk socjologicznych, prezentowanych przykładowo w pierwszym podrozdziale rozdziału I niniejszej pracy.

Europejska droga do regulacji zakazu dyskryminacji w jego obecnym kształcie opierała się częściowo na doświadczeniach postkolonialnych, a w pozostałym zakresie na doświadczeniach II wojny światowej. Skutkiem dojścia do głosu przez grupy historycznie marginalizowane były reorganizacja i regulacja praw człowieka oraz społeczna presja na władzę w celu promowania równości i naturalnej sprawiedliwości.

Wraz z rozwojem międzynarodowego prawa podmiotowego państwa na mocy Powszechnej Deklaracji Praw Człowieka[244] zobowiązały się do przeciwdziałania bezpośredniej i pośredniej dyskryminacji. W tamtym okresie rola Wspólnot Europejskich (obecnie Unii Europejskiej) w tworzeniu prawa antydyskryminacyjnego była jednak znikoma. Wspólnoty jednoczyły swoich członków głównie w obszarze gospodarczym, a gwarancje prawnoczłowiecze zawarto dopiero w Europejskiej konwencji o ochronie praw człowieka i podstawowych wolności z 1950 r. Ogólny zakaz dyskryminacji, wprowadzony art. 14 Konwencji[245], wyznaczył oś standardu równości w systemie prawnym Rady Europy[246]. Równość i promocja praw człowieka po raz

[243] Krajowy ustawodawca nie zdecydował się jak dotychczas na wprowadzenie ujednoliconych przepisów antydyskryminacyjnych, chociaż taką praktykę można spotkać w pozostałych krajach członkowskich UE.
[244] Powszechna Deklaracja Praw Człowieka, Paryż, 10.12.1948 r., http://libr.sejm.gov.pl/tek01/txt/onz/1948.html (dostęp: 4.12.2021); dalej: PDPCz.
[245] Charakter prawny zakazu dyskryminacji zawartego w art. 14 EKPCz jest sporny. W piśmiennictwie wyrażane są poglądy, że artykuł ten nie przyznaje samoistnego prawa lub wolności analogicznego do tych, które Konwencja gwarantuje w przepisach poprzedzających (art. 1–13), lecz podobnie jak art. 15–18 odnosi się do zakresu i wykonywania praw i wolności gwarantowanych Konwencją; por. J.G. Merrills, A.H. Robertson, *Human Rights in Europe*, New York-Manchester, UK 2001, s. 199.
[246] M. Kułak, K. Śmiszek, *Kilka uwag o standardzie równości w Europejskiej Konwencji Praw Człowieka*, „EPS" 2, 2017, s. 40–45.

76:36584

pierwszy w historii stały się przedmiotem aktów UE w traktacie rzymskim z 1957 r.[247] jako filary integracji politycznej państw Europy.

W okresie drugiej fali feminizmu przełomowe okazało się orzecznictwo Trybunału Sprawiedliwości Unii Europejskiej (dalej: TSUE). Społeczna niezgoda manifestująca się w fizycznych oznakach sprzeciwu wobec rażących nierówności w dystrybuowaniu dóbr oraz dostępie do praw i wolności pomiędzy płciami doprowadziły do pozytywizacji praw wyborczych kobiet i sformalizowania równości płci[248]. Na skutek wskazanych zdarzeń Europejska Wspólnota Gospodarcza uchwaliła dyrektywę o równości wynagrodzeń[249], postulującą równe płace dla obu płci, a następnie dyrektywę o równym traktowaniu[250], która zakazywała dyskryminacji ze względu na płeć w miejscu pracy oraz przy dostępie do szkoleń zawodowych i edukacji.

Z czasem regulacje równościowe obejmowały nowe sektory. Zmiany przyspieszały również wskutek propagowania przez TSUE postulatów równościowych, m.in. dzięki rozwinięciu koncepcji dyskryminacji pośredniej. Trybunał, odpowiadając na zmieniające się formy różnicowania, doprowadził wówczas do wykształcenia efektywniejszego narzędzia ochrony w przypadku niedozwolonego różnicowania, które nie opierało się na jawnym różnicowaniu ze względu na cechy chronione. Zakaz niesprawiedliwego różnicowania nadal w sposób ograniczony regulował jednak stosunki społeczne, wpływając na poszczególne sfery korzystania z praw i wolności oraz ograniczając się do ochrony różnicowania ze względu na określone cechy.

Kolejnym etapem wykształcania się regulacji było poszerzanie katalogu cech chronionych o inne niż płeć i narodowość. Kluczowy w tym zakresie był traktat amsterdamski z 1997 r.[251], który zakazał różnicowania ze względu na rasę, wiek, pochodzenie etniczne, orientację seksualną, niepełnosprawność czy religię lub przekonania[252]. Standard równości został również

[247] Włączenie do Europejskiego Traktatu Gospodarczego (inaczej: traktat rzymski) z 1957 r. (dalej: EWG) przesłanek dyskryminacji było ukoronowaniem przekonania, że Europa powinna być jednolitym rynkiem gospodarczym.
[248] Doktryna równości płci została rozwinięta na podstawie art. 119 EWG i stała się jedną z podstawowych zasad prawa UE.
[249] Dyrektywa Rady z 10.02.1975 r. w sprawie zbliżenia ustawodawstw Państw Członkowskich dotyczących stosowania zasady równości wynagrodzeń dla mężczyzn i kobiet, Dz.Urz. L 045 z 19.02.1975 r., s. 0019–0020.
[250] Dyrektywa Rady z 9.02.1976 r. w sprawie wprowadzenia w życie zasady równego traktowania kobiet i mężczyzn w zakresie dostępu do zatrudnienia, kształcenia i awansu zawodowego oraz warunków pracy, Dz.U. L 39 z 14.02.1976 r., s. 40–42.
[251] Traktat z Amsterdamu, Dz.U. C 340 z 10.11.1997 r., 11997D/TXT.
[252] Art. 13 traktatu amsterdamskiego.

77:88885

wzmocniony przez jednoznaczne wprowadzenie do celów Wspólnoty Europejskiej[253] dążenia do osiągnięcia równości kobiet i mężczyzn.

W ślad za poszerzeniem katalogu cech chronionych traktatem amsterdamskim wprowadzono dyrektywę o równości rasowej[254] oraz dyrektywę ramową w sprawie równości zatrudnienia[255], która jako pierwszy akt prawa wtórnego zakresem przedmiotowym obejmowała także „dostęp do dóbr i usług oraz dostarczania dóbr i usług publicznie dostępnych, włącznie z zakwaterowaniem"[256]. Dyrektywy UE o niedyskryminacji nie objęły jednak zakazem dyskryminacji dotyczącej statusu finansowego.

Następnym krokiem w kierunku realizacji postulatów równościowych było nałożenie na państwa członkowskie na mocy traktatu lizbońskiego[257] obowiązku aktywnego zwalczania dyskryminacji. Przyjęcie traktatu zrównuje się w doktrynie europejskiego prawa antydyskryminacyjnego z wprowadzeniem nowego wymiaru chronionych kryteriów w prawie pierwotnym, które nie mogą być podstawą nieusprawiedliwionego nierównego traktowania[258]. Ten nowy wymiar przejawiał się w nadaniu wiążącego charakteru prawnego Karcie Praw Podstawowych Unii Europejskiej, w wyniku którego stała się ona elementem prawa pierwotnego UE. W ten sposób Unia, mimo początkowych postulatów gospodarczych, przeszła drogę intensywnej integracji aż do przyjęcia katalogu gwarantowanych wolności i praw człowieka o charakterze konstytucyjnym. Dzięki temu europejskie prawo antydyskryminacyjne zyskało charakter prawa podmiotowego publicznego[259] o wymiarze konstytucyjnym[260], którego przedmiot, zakres i treść określają przepisy prawa pierwotnego UE.

Ze względu na podstawę prawną, zakres podmiotowy, zakres normowania i zastosowania, strukturę formalną i charakter prawny, ograniczenia, a także uzasadnienie oraz funkcje

[253] Art. 2 oraz art. 3 ust. 2 traktatu amsterdamskiego.
[254] Dyrektywa Rady 2000/43/WE z 29.06.2000 r. wprowadzająca w życie zasadę równego traktowania osób bez względu na pochodzenie rasowe lub etniczne, Dz.U. L 180 z 19.07.2000 r., s. 22–26.
[255] Dyrektywa Rady 2000/78/WE z 27.11.2000 r. ustanawiająca ogólne warunki ramowe równego traktowania w zakresie zatrudnienia i pracy, Dz.U. L 303 z 2.12.2000 r., s. 16–22.
[256] Art. 3 ust. 1 pkt h) dyrektywy Rady 2000/78/WE.
[257] Traktat z Lizbony zmieniający Traktat o Unii Europejskiej i Traktat ustanawiający Wspólnotę Europejską podpisany w Lizbonie 13.12.2007 r., Dz.Urz. UE C 306 z 17.12.2007 r., 12007L/TXT.
[258] D. Schiek, A. Lawson, *European Union Non-Discrimination Law and Intersectionality, Investigating the Triangle of Racial, Gender and Disability Discrimination*, London 2011, s. 12.
[259] Przepis prawny dający osobie zainteresowanej możność domagania się określonego zachowania władzy, stwarzający dla tej osoby prawo podmiotowe publiczne. Więcej zob. B. Adamiak, J. Borkowski, *Kodeks postępowania administracyjnego. Komentarz*, Warszawa 2021.
[260] Art. 9 Konstytucji stanowi, że Rzeczpospolita Polska będzie przestrzegać wiążącego ją prawa międzynarodowego, a zatem również KPP i EKPCz. Krytycznie przy tym należy odnieść się do wyroku polskiego Trybunału Konstytucyjnego z 7.10.2021 r. (sygn. K 3/21), w którym przyjęto, że prawo unijne nie ma prymatu nad polskimi przepisami konstytucyjnymi.

78:80142

europejskie zakazy dyskryminacji i korespondujące z nimi podmiotowe prawa publiczne do niedyskryminacyjnego traktowania tworzą heterogeniczną grupę praw podmiotowych[261]. Prawa podstawowe są integralną częścią zasad ogólnych prawa[262], w związku z czym niedopuszczalne jest podejmowanie przez UE środków, które byłyby sprzeczne z poszanowaniem tych praw[263]. Tym samym postulaty równościowe powinny nie tylko przejawiać się w działaniach władzy na poziomie regulacji stosunków publicznoprawnych czy prywatnoprawnych jednostki, ale również być realizowane i rozwijane w dalszych aktach prawnych Unii. W tym sensie zakaz dyskryminacji może być postrzegany jako ograniczenie podstawowych swobód rynku wewnętrznego[264].

W nowożytnym europejskim prawie antydyskryminacyjnym wyróżniamy dwie podstawowe formy dyskryminacji, omówione szczegółowo w rozdziale I niniejszej pracy, tj. dyskryminację bezpośrednią i pośrednią. Dyskryminacja bezpośrednia jest zdefiniowana w prawie UE jako sytuacja, w której adresat decyzji jest traktowany mniej korzystnie niż jest, był lub byłby traktowany inny podmiot w porównywalnej sytuacji[265]. Na tym gruncie art. 21 KPP zakazuje wszelkiej dyskryminacji, zwłaszcza ze względu na płeć, rasę, kolor skóry, pochodzenie etniczne lub społeczne, cechy genetyczne, język, religię albo przekonania, poglądy polityczne lub wszelkie inne poglądy, przynależność do mniejszości narodowej, majątek, urodzenie, niepełnosprawność, wiek bądź orientację seksualną. Co więcej, w zakresie zastosowania traktatów i bez uszczerbku dla ich postanowień szczególnych zakazana jest wszelka dyskryminacja ze względu na przynależność państwową.

Zgodnie zaś z art. 14 EKPCz korzystanie z praw i wolności wymienionych w Konwencji powinno być zapewnione bez dyskryminacji wynikającej z takich cech jak płeć, rasa, kolor skóry, język, religia, przekonania polityczne i inne, pochodzenie narodowe lub społeczne, przynależność do mniejszości narodowej, majątek, urodzenie bądź z jakichkolwiek innych przyczyn. Regulację tę uzupełnia protokół dodatkowy nr 12 do EKPCz, zgodnie z którym nikt

[261] J. Maliszewska-Nienartowicz, *Prawo antydyskryminacyjne*…, nb. 78.
[262] Zgodnie z art. 6 traktatem o Unii Europejskiej KPP ma ten sam status prawny co traktaty, ale nie jest ich elementem.
[263] Wyrok TSUE z 12.06.2003 r. w sprawie *Eugen Schmidberger, Internationale Transporte und Planzüge v. Republik Österreich*, skarga nr C-112/00, pkt 62, 64.
[264] TSUE nie ustanowił dotychczas sztywnej zasady pozwalającej na ustanowienie hierarchii między prawami podstawowymi a podstawowymi swobodami; por. wyrok TSUE z 12.06.2003 r. w sprawie *Eugen Schmidberger*…
[265] Por. art. 2(2)(a) dyrektywy Rady 2000/43/WE.

79:87607

nie może być dyskryminowany przez jakiekolwiek władze publiczne z któregokolwiek z wymienionych powodów.

Jak już wspomniałam, koncepcję dyskryminacji pośredniej wykształcił Trybunał Sprawiedliwości Unii Europejskiej w połowie lat 70. XX w. w odniesieniu do narodowości i płci[266]. Pod tą formą dyskryminacji rozumiemy proces, w którym pozornie neutralny przepis, kryterium lub praktyka stawia członków kategorii chronionej w szczególnie niekorzystnej sytuacji w porównaniu z innymi podmiotami, chyba że taki przepis, kryterium lub praktyka jest obiektywnie uzasadniony zgodnym z prawem celem, a środki służące jego osiągnięciu są właściwe i odpowiednie[267].

Mając na względzie zarówno przybliżone drogi do objęcia niesprawiedliwego różnicowania zakazem dyskryminacji, jak i historyczne uwarunkowania wpływające na rodzaje form dyskryminacji, należy wskazać na zróżnicowanie problemu, na który miała odpowiadać tak wykształcona regulacja. Oparcie prawnego zakazu dyskryminacji na określonych cechach chronionych i widocznym oraz uzasadnialnym procesie różnicowania pomiędzy możliwymi do zidentyfikowania grupami wobec niejawnego i nieprzejrzystego procesu decyzyjnego ASD, posadowionego zasadniczo na więcej niż jednym kryterium (niekoniecznie spełniającym warunki cechy chronionej), budzi uzasadnione obawy o możliwość realizowania postulatów równościowych przy wykorzystaniu istniejących, historycznie wykształconych narzędzi prawa antydyskryminacyjnego. To zaś uzasadnia refleksję nad adekwatnością przepisów powstałych na gruncie historycznie ustalonych cech chronionych i stosunków różnicowania, w odniesieniu do których władza na skutek walki o równość sama narzucała sobie pewne rozwiązania, za którymi podążało budowanie świadomości społecznej. Zmiana krajobrazu stosunków dominacji może wpływać na to, że zakaz dyskryminacji przestanie ograniczać rzeczywisty problem nierówności. Można bowiem wskazać, iż – podobnie jak w przypadku wykształcenia się formy dyskryminacji pośredniej – aktualnie następuje taka modyfikacja stosunków społecznych, która może spowodować ewolucję myślenia równościowego.

[266] Dyskryminacja pośrednia ze względu na płeć została uznana w przypadku mniej korzystnego traktowania pracowników zatrudnionych w niepełnym wymiarze czasu pracy. Pośrednio dotyczyło to zwłaszcza kobiet, które ze względu na społeczne role związane z rodzicielstwem stanowiły większość pracowników zatrudnionych w niepełnym wymiarze czasu pracy; por. D. Morondo Taramundi, *op. cit.*, s. 77.
[267] Por. art. 2(2)(b) dyrektywy Rady 2000/43/WE.

80:43122

2. Europejskie prawo antydyskryminacyjne jako podstawa zakazu różnicowania przez ASD w świetle orzecznictwa sądów

Analizując historyczną walkę o równość w nowożytnych państwach demokratycznych, można zaobserwować, że prawo antydyskryminacyjne jest niekompletnym, ale dynamicznie rozwijanym mechanizmem ograniczania historycznych niesprawiedliwości i głęboko zakorzenionych struktur dominacji oraz wykluczenia. W toku wykształcania się zakazu dyskryminacji widoczna jest przy tym stała progresja przepisów, a także obszarowość i rozproszenie regulacji związane z tłem społeczno-politycznym urealniania postulatów równościowych. Tak powstałe zbiory miały status wyodrębnionych części i mogły względnie samodzielnie funkcjonować w danych obszarach zakazu nieuzasadnionego różnicowania. Fragmenty te wzmacniały jednocześnie teorie równości materialnej i szersze postulaty społeczne.

Na skutek tego zasada niedyskryminacji doprowadziła do ukształtowania ogólnego zakazu dyskryminacji i stała się nieodłącznym elementem międzynarodowego systemu ochrony praw człowieka, realizowanego m.in. w Powszechnej Deklaracji Praw Człowieka[268], Europejskiej konwencji praw człowieka i podstawowych wolności, Traktacie o funkcjonowaniu Unii Europejskiej[269] czy Karcie Praw Podstawowych Unii Europejskiej[270]. Wskazane nowożytne regulacje wzajemnie się przenikają, tworząc przedmiotowy i podmiotowy zakres zakazu dyskryminacji. Przejawem tego uniwersalnego dążenia jest wprowadzenie w Traktacie o funkcjonowaniu UE wspólnych wartości i celów regulacji europejskich, m.in. zwalczania wszelkiej dyskryminacji na podstawie idei poszanowania praw człowieka[271].

Na tak wykształcony system ogólnego prawa antydyskryminacyjnego składają się nie tylko przepisy prawa międzynarodowego oraz europejskiego, ale również prawa krajowego, które

[268] Art. 2 PDPCz stanowi, że każdy człowiek posiada wszystkie prawa i wolności zawarte w niniejszej deklaracji bez względu na jakiekolwiek różnice rasy, koloru, płci, języka, wyznania, poglądów politycznych i innych, narodowości, pochodzenia społecznego, majątku, urodzenia lub jakiegokolwiek innego stanu. Nie wolno ponadto czynić żadnej różnicy w zależności od sytuacji politycznej, prawnej lub międzynarodowej kraju bądź obszaru, do którego dana osoba przynależy, bez względu na to, czy dany kraj lub obszar jest niepodległy, czy też podlega systemowi powiernictwa, nie rządzi się samodzielnie lub jest w jakikolwiek sposób ograniczony w swej niepodległości.

[269] Art. 19 Traktatu o funkcjonowaniu Unii Europejskiej (dalej: TFUE) stanowi, że bez uszczerbku dla innych postanowień traktatów i w granicach kompetencji, które traktaty powierzają Unii, Rada, stanowiąc jednomyślnie zgodnie ze specjalną procedurą ustawodawczą i po uzyskaniu zgody Parlamentu Europejskiego, może podjąć środki niezbędne do zwalczania wszelkiej dyskryminacji ze względu na płeć, rasę lub pochodzenie etniczne, religię lub światopogląd, niepełnosprawność, wiek lub orientację seksualną.

[270] Rozdział III Karty Praw Podstawowych Unii Europejskiej pt. *Równość*.

[271] J. Maliszewska-Nienartowicz, *Karta Praw Podstawowych*…, preprint 20.

81:76077

zazwyczaj podąża za celami określanymi ponadnarodowo[272]. Relewantną rolę odgrywa też dotychczasowe orzecznictwo sądowe Trybunału Sprawiedliwości Unii Europejskiej oraz Europejskiego Trybunału Praw Człowieka, jako że sądownictwo stale adaptuje i uaktualnia rolę zakazu dyskryminacji na tle społecznych nierówności i stosunków dominacji, dzięki czemu przepisy zachowują wymiar adekwatności do problemu dyskryminacji.

Można przy tym wskazać na dwa kluczowe aspekty związane z ochroną praw podstawowych w kontekście zabezpieczenia na drodze wykładni prawa przez orzecznictwo podstawowych swobód rynku wewnętrznego, mianowicie aspekt praktyczny i teoretyczny[273]. Aspekt praktyczny można ocenić pozytywnie, gdyż aktualnie prawa podstawowe są chronione poprzez założenie, że w sposób uzasadniony w toku oceny *casu ad casum* mogą one ograniczać podstawowe swobody gospodarcze. Jak jednak wskazuje Joanna Ryszka[274], z teoretycznego punktu widzenia sprawa wygląda o wiele gorzej. Pojawiają się bowiem trudności w zakresie zdefiniowania statusu tych praw. W kontekście niniejszej analizy ma to istotne znaczenie dla ograniczenia wolności gospodarczej poprzez uregulowanie granic algorytmicznej predykcji, umożliwiającej optymalizację procesów decyzyjnych. To zaś nie służy poczuciu pewności i bezpieczeństwa prawnego.

Nie zawsze również stanowisko TSUE odpowiadało oczekiwaniom marginalizowanych społeczności[275]. W związku z tym uznaję, że dalszą analizę prawną regulacji szczegółowych należy poprzedzić zbadaniem dorobku orzeczniczego mogącego wprost lub przez analogię znaleźć zastosowanie do ograniczania wpływu automatyzacji na realizację postulatów równościowych.

W kontekście algorytmów dyskryminacja bezpośrednia obejmuje przede wszystkim sytuacje, w których model ASD różnicuje w zakresie reguły lub kodu na podstawie chronionej cechy. Podstawowym elementem definicji zakazu dyskryminacji, który jest fundamentalny z punktu

[272] Są to: Traktat o funkcjonowaniu Unii Europejskiej, Karta Praw Podstawowych Europejskiej konwencji o ochronie praw człowieka i podstawowych wolności (wraz z protokołem dodatkowym nr 12 do EKPCz), Konwencja o ochronie praw człowieka i podstawowych wolności, Konwencja ramowa o ochronie mniejszości narodowych z 1.02.1995 r. oraz Konwencja ONZ o prawach osób niepełnosprawnych.

[273] S. Robin-Olivier, *Protection of Fundamental Rights under Internal Market Rules and in the Area of Freedom, Security and Justice: Unity or Dualism?*, University of Paris Quest-Naterre La Défense – Working Paper, Paris 2009, s. 5.

[274] J. Ryszka, *Prawa podstawowe jako uzasadnione ograniczenie swobód rynku wewnętrznego Unii Europejskiej*, „Przegląd Prawa i Administracji" 96, 2014, s. 59–78.

[275] TSUE nie zawsze skutecznie realizował założenia prawa antydyskryminacyjnego. Jako przykład można podać wyrok w sprawie *Geoffrey Léger v. Ministrie des Affaires sociales* (skarga nr C-528/13), w którym Trybunał uznał, że stałe wykluczenie mężczyzn mających stosunki seksualne z innymi mężczyznami z oddawania krwi może – w związku z sytuacją panującą w danym państwie członkowskim – być uzasadnione.

82:67548

widzenia hipotezy przedmiotowej niniejszej pracy, jest zatem niedozwolone kryterium, obecnie definiowane jako cecha chroniona.

Doktryna prawa antydyskryminacyjnego dopuszcza przynajmniej trzy modele formułowania niedozwolonych kryteriów różnicujących: 1. zamknięty katalog cech chronionych opisany przepisami prawa; 2. otwarty katalog cech chronionych; 3. brak katalogu cech zastąpiony ogólną zasadą równości[276]. Według Sandry Fredman charakterystyczne dla pierwszego modelu jest nowelizowanie i uaktualnianie właściwych przepisów prawa w celu dostosowania zakazu do zmieniającej się rzeczywistości społeczno-politycznej (na takim modelu oparta jest chociażby konstrukcja przepisu art. 19 ust. 1 TFUE). Drugi model zakłada aktywną rolę sądów w uzupełnianiu nazwanych i wymienionych w przepisach prawa kryteriów różnicujących (jak np. w przypadku art. 14 EKPCz lub art. 21 KPP). W ostatnim – trzecim – modelu orzecznictwo sądowe odgrywa rozstrzygającą rolę w konstruowaniu nie tylko ram równości społecznej, lecz także treści zakazu dyskryminacji, będącego wynikową oceny *ad casum* poszczególnych spraw dotyczących naruszenia zasady równości.

Ocena adekwatności modeli formułowania katalogów cech powinna jednak zostać poprzedzona analizą realnej możliwości wskazania tych cech w sposób prawidłowy i zupełny. Jak już wspomniałam, aktualnie przepisy Karty Praw Podstawowych zakazują dyskryminacji, zwłaszcza ze względu na płeć, rasę, kolor skóry, pochodzenie etniczne lub społeczne, cechy genetyczne, język, religię albo przekonania, poglądy polityczne bądź wszelkie inne poglądy, przynależność do mniejszości narodowej, majątek, urodzenie, niepełnosprawność, wiek lub orientację seksualną, a w zakresie zastosowania traktatów również przynależność państwową[277]. Europejska konwencja praw człowieka i podstawowych wolności wymienia zaś takie cechy jak płeć, rasa, kolor skóry, język, religia, przekonania polityczne i inne, pochodzenie narodowe lub społeczne, przynależność do mniejszości narodowej, majątek i urodzenie[278]. Oba katalogi są otwarte, przy czym w literaturze przedmiotu uznaje się za mało prawdopodobne, aby realne było uzupełnienie tych cech o nowe kategorie[279].

[276] S. Fredman, *Discrimination Law*, Oxford 2002, s. 113.
[277] Art. 21 KPP.
[278] Art. 14 EKPCz.
[279] Autorka wnioskuje z analizy orzecznictwa TSUE, wskazując, że Trybunał nie był skłonny do przyznawania ochrony na podstawie takich przesłanek jak przewlekła choroba, otyłość lub przypadki, w których krzyżują się seksualność i wiek, nawet jeśli wyraźnie występuje niekorzystna sytuacja społeczna. Trybunał poszukiwał w tych przypadkach łącznika do cech wprost chronionych; por. S. Wachter, *op. cit.*, s. 14.

83:58927

Wyróżnienie cechy chronionej w toku różnicowania algorytmicznego jest niezwykle utrudnione, a w wielu przypadkach niemożliwe. ASD może więc prowadzić do wykluczenia i marginalizacji na podstawie cech nieuznanych za chronione prawem antydyskryminacyjnym, informacji nieobserwowalnych lub aspektów behawioralnych, a nadto zestawów tych danych i cech. Tego rodzaju kategoryzowanie ludzi odbiega od tradycyjnie rozumianych procesów różnicowania, a dodatkowo jest – w wypadku zaawansowanych algorytmów uczenia maszynowego – niewytłumaczalne nawet dla twórcy czy operatora systemu. Problematyka ta uwidoczniła się już w prezentowanych przykładach błędnych ASD, szczególnie niderlandzkiego SyRI i polskiego systemu do profilowania bezrobotnych, które opierały swoje rozstrzygnięcia na wartościowaniu (bliżej nieujawnionych) zestawów cech, danych i informacji o jednostce. W konsekwencji formułowanie zamkniętych katalogów wydaje się niemożliwe w odniesieniu do specyfiki różnicowania przez ASD, co więcej – pojawiają się dalej idące wątpliwości, czy analiza bezprawności takiego procesu powinna się skupiać na procesie wartościowania przez algorytm[280].

Dotychczasowe debaty prawne nad sprawiedliwością algorytmiczną były zakorzenione w amerykańskim prawie antydyskryminacyjnym przez co opierały się na wykształconych na gruncie tego prawa kategoriach różnicującego traktowania i jego wpływu na sytuację jednostki[281]. Z tej przyczyny przyjmowano, że wobec niemożliwości przypisania intencji dyskryminacji i dostatecznych narzędzi związanych z wykazaniem różnicowania przez ASD na podstawie cech skatalogowanych w aktach prawa dyskryminacja automatyczna będzie miała charakter pośredni. Ta refleksja początkowo ograniczyła europejską debatę, uniemożliwiając rozpoczęcie badań nad zakwalifikowaniem dyskryminacji automatycznej jako bezpośredniej. Tymczasem na gruncie prawa europejskiego i międzynarodowego niezamierzona dyskryminacja może być bezpośrednia, dzięki czemu sama regulacja mogłaby znaleźć lepsze zastosowanie. To jednak wymagałoby uprzedniego redefiniowania zastosowania normy z perspektywy sposobu działania ASD.

Pierwszym problemem w tym zakresie jest trudność detekcji dyskryminacji bezpośredniej w ASD wobec niejawnych, nieprzejrzystych i niewidocznych procesów różnicowania. Pokrzywdzony różnicowaniem nie ma bowiem odpowiednich (a przede wszystkim

[280] Por. R. Xenidis, L. Senden, *EU Non-Discrimination Law in the Era of Artificial Intelligence: Mapping the Challenges of Algorithmic Discrimination*, [w:] *General Principles of EU Law and the EU Digital Order*, red. U. Bernitz *et al*., Alphen aan den Rijn 2020, s. 151–181.
[281] J. Adams-Prassl, R. Binns, A. Kelly-Lyth, *Directly Discriminatory Algorithms*, „Modern Law Review" 86, 2022, nr 1, s. 174.

84:70025

skutecznych) narzędzi umożliwiających mu stwierdzenie faktów. W sprawie *Galina Meister v. Speech Design Carrier Systems GmbH*[282] TSUE uznał, że prawo do informacji o zatrudnieniu nie istnieje w stosunku do pracodawcy, nawet jeśli odmowa udzielenia takich informacji przez przedsiębiorstwo może być elementem służący do stwierdzenia dyskryminacji *prima facie*. Trybunał oparł się w tym wypadku na nieistnieniu prawa jednostki do realnego zbadania procesu decyzyjnego w celu potwierdzenia jego dyskryminującego charakteru, wskazując jednocześnie, że to na pokrzywdzonym spoczywa obowiązek przedstawienia faktów, z których można domniemywać wystąpienie dyskryminacji. Jedynie w przypadku, gdy taka osoba przedstawi odpowiednie fakty, do naruszyciela będzie należeć udowodnienie, że nie miało miejsca naruszenie zasady niedyskryminacji. Na tle wskazanej sprawy uwidocznia się nie tylko problem ustalenia istnienia różnicowania ze względu na cechę chronioną, ale też niemożliwości przeprowadzenia porównania, zarówno ze względu na trudności w identyfikacji grupy porównywalnej, jak i wykazanie faktu potraktowania grup w sposób zróżnicowany.

W wyroku TSUE pominął postulat przejrzystości jako warunek wstępny do skutecznego egzekwowania zasady niedyskryminacji. To niezwykle istotne, albowiem dostęp do wiedzy o funkcjonowaniu algorytmów jest ograniczony zarówno dla operatora systemu i adresata decyzji, jak i nawet samego twórcy ASD. Nie tylko wynika to z kwestii technologicznych związanych z systemami czarnoskrzynkowymi czy uczenia algorytmu na danych, ale również jest powodowane motywami rynkowymi. Odmowa ujawnienia informacji jest bowiem korzystna dla naruszyciela z wielu perspektyw i powszechnie wykorzystywana. Sformułowało ją nawet Królestwo Niderlandów w przypadku systemu SyRI, powołując się na względy bezpieczeństwa. Co więcej, bez ingerencji prawodawcy taka tendencja będzie się jedynie pogłębiała.

Z opisanych wcześniej przyczyn badania nad możliwością penalizacji algorytmicznego różnicowania koncentrują się na skutkach procesu decyzyjnego, prowadząc do kwalifikowania tego różnicowania jako pośredniej formy dyskryminacji[283]. Dyskryminacja pośrednia wydaje się bowiem lepiej odpowiadać na problem niejawnego i nieprzejrzystego procesu decyzyjnego, gdyż zrozumienie samego procesu w tym wypadku nie jest konieczne, wykorzystywane zaś w nim kryterium może uchodzić za neutralne.

[282] Wyrok TSUE (druga izba) z 19.04. 2012 r. w sprawie *Galina Meister v. Speech Design Carrier Systems GmbH*, skarga nr C-415/10.
[283] R. Xenidis, L. Senden, *op. cit*.

85:57282

Jak wskazują Raphaële Xenidis i Linda Senden, ofiary dyskryminacji, ale również orzekające sądy nie mają przy tym potrzeby znać dokładnie sposobu działania bezprawnie różnicującego algorytmu[284]. Do dyskryminacji pośredniej będzie jednak dochodzić wyłącznie wtedy, gdy odmienne traktowanie będzie się odbywać bez obiektywnego i racjonalnego uzasadnienia różnicowania osób zaliczanych do grup chronionych znajdujących się w podobnych sytuacjach. Andrzej Wróbel wskazuje co prawda, że podmiot nie jest dyskryminowany ze względu na przynależność do grupy osób mających daną cechę/atrybut, lecz ze uwagi na to, że to on dysponuje taką cechą/atrybutem[285], aczkolwiek nadal konieczna jest przy tym wiedza o konkretnej przyczynie różnicowania.

W wypadku automatycznych procesów decyzyjnych profile budowane na podstawie danych tworzą sztuczne i niereprezentatywne kategorie, przez co trudno sprowadzić je do rzeczywistych, realnie istniejących grup społecznych[286]. Rzadko grupy te będą identyfikowane na podstawie jednej chronionej cechy, albowiem zestawy danych zazwyczaj będą prowadziły do krzyżowania się poszczególnych kategorii. W orzeczeniu w sprawie *Parris v. Trinity College Dublin* TSUE z jednej strony uznał wystąpienie dyskryminacji wielokrotnej, stwierdzając, że dyskryminacja może rzeczywiście opierać się na cechach chronionych prawem UE, ale z drugiej – odrzucił stwierdzenie dyskryminacji intersekcjonalnej, oświadczając, że nie można stwierdzić żadnej nowej kategorii dyskryminacji wynikającej z połączenia więcej niż jednej z tych podstaw, jeśli nie stwierdzono dyskryminacji ze względu na te przyczyny rozpatrywane oddzielnie[287].

Możliwe jest również takie algorytmiczne kategoryzowanie, którego przyczyn ludzie nie będą rozumieli. W tej sytuacji jednostka objęta procesem różnicowania nie będzie mogła zauważyć, czy i kiedy staje się częścią grupy, a także nie będzie potrafiła samodzielnie wskazać innych podmiotów zaliczanych do takiej grupy. Takie procesy różnicowania – które mogą jednocześnie mieć skutek dyskryminacyjny – będą odbiegać od historycznie wykształconych form jawnej stygmatyzacji. Mimo to – z powodów opisanych w rozdziale II – negatywne skutki stosowania ASD będą odczuwalne szczególnie wśród grup marginalizowanych, które w szerszej perspektywie są ofiarami procesów wykluczenia. Pokrzywdzeni niesprawiedliwym różnicowaniem będą w konsekwencji – ze względów majątkowych, społecznych i

[284] *Ibidem*.
[285] Zob. J. Maliszewska-Nienartowicz, *Karta Praw Podstawowych*…, preprint 21, par. 16.
[286] M. Leese, *op. cit.*, s. 504.
[287] J. Gerards, R. Xenidis, *op. cit.*; zob. wyrok TSUE z 24.11.2016 r. w sprawie *David L. Parris v. Trinity College Dublin i in.*, skarga nr C-443/15.

86:98023

politycznych – ograniczeni w dostępie do narzędzi prawnych i odpowiedniej reprezentacji potrzebnej, aby samodzielnie przeciwstawić się pogłębiającym się procesom marginalizacji na skutek stosowania ASD.

Tym samym istnieje prawdopodobieństwo zwiększania zasięgu bezprawnego różnicowania i skali problemu społecznego. Wpływa na to nie tylko masowość automatycznych systemów decyzyjnych, ale również skala i tempo utrwalania się uprzedzeń poprzez uczenie ASD wzorców z obciążonych historycznie uprzedzeniami zbiorów danych. Identyfikacja i porównanie grup, które formułowane są dynamicznie przez ASD na podstawie analizy zespołu cech i danych, jak w przypadku SyRI czy polskiego ASD do profilowania bezrobotnych, wydają się problematyczne nawet dla użytkownika publicznoprawnego czy dostawców systemu.

Trudności w identyfikacji grup i zróżnicowany zakres ochrony przed kryteriami dyskryminującymi są potęgowane uprawnieniem do uzasadnienia nierównego traktowania[288]. Przyjmuje się w tym zakresie, że brak obiektywnego i racjonalnego uzasadnienia występuje, gdy rozpatrywane zróżnicowanie nie służy realizacji legitymowanego celu bądź też gdy nie zachodzi racjonalny stosunek proporcjonalności między zastosowanymi środkami a celem, który mają one zrealizować[289].

Z zakwalifikowaniem różnicowania jako dyskryminacji pośredniej łączy się konieczność przeprowadzenia testu proporcjonalności[290]. Na pierwszym etapie należy wykazać zgodność celu różnicowania z prawem. To badanie nie jest krytyczne dla zakazu, albowiem ASD nie są zasadniczo stosowane w celach bezprawnych – celem automatyzacji procesów decyzyjnych jest wszak racjonalizacja kosztów przy jednoczesnej bardzo dokładnej analizie danych na podstawie zaobserwowanych zależności. Dyskryminacja ta będzie zatem w dużej mierze realizowała się w sposób podobny do dyskryminacji statystycznej. Jej skutkiem co do zasady nie będzie jednak wprost bezprawność. Niemniej odkrywanie i utrwalanie pewnych korelacji z zaobserwowanych wzorców może mieć oczywiście niekorzystne społecznie skutki (np. umacnianie modelu, w którym kobiety zarabiają mniej od mężczyzn).

[288] M. Domańska, *Zakaz dyskryminacji ze względu na więcej niż jedno zabronione kryterium*, Warszawa 2019, rozdz. 1.4.1.
[289] Zob. wyrok ETPCz z 22.12.2009 r. w sprawie *Sejdić i Finci*… (par. 42), stanowiący najnowszą wypowiedź Wielkiej Izby i potwierdzający definicje przyjęte we wcześniejszym orzecznictwie.
[290] D. Morondo Taramundi, *op. cit.*, s. 80.

87:86247

Warunek braku bezprawności musi być w związku z tym uzupełniony badaniem proporcjonalności różnicowania. Jeżeli więc istnieje inne narzędzie służące osiągnięciu tego samego celu co ASD, które w mniejszym stopniu naruszałoby prawo antydyskryminacyjne, to rozwiązanie algorytmiczne nie byłoby zgodne z prawem[291]. Ustalenie proporcjonalności będzie się zatem sprowadzało do przeprowadzenia przez sądy analizy balansu między precyzją, efektywnością i szybkością ASD a skutkami uczenia maszynowego mogącego prowadzić do powielania zaobserwowanych wzorców na masową skalę.

Test proporcjonalności zazwyczaj pozostawiany jest przez TSUE do przeprowadzenia sądom krajowym, ponieważ wymaga ustalenia i oceny faktów leżących u podstaw danej sprawy[292]. Ochrona przed dyskryminującym efektem ASD doznaje przez to daleko idących ograniczeń. Sądy krajowe mogą bowiem nieobiektywnie lub niechętnie kontrolować predykcyjne narzędzia algorytmiczne (w tym stosowane w administracji państwowej), szczególnie jeżeli będą one pomocne w kwestiach bezpieczeństwa publicznego czy dystrybucji środków z obszaru zabezpieczeń społecznych. Jednostka nie ma przy tym odpowiednich indywidualnych narzędzi do weryfikacji ASD, co więcej – wielokrotnie nie jest świadoma wykorzystywania systemów automatycznych do rozstrzygania o jej prawach i obowiązkach. Potwierdzają to wspominane już wielokrotnie przypadki SyRI i polskiego ASD do profilowania bezrobotnych.

Dodatkowym problemem będzie fragmentaryczność oraz rozdrobnienie procesu tworzenia i implementacji algorytmicznych systemów decyzyjnych, a także wrażenie pozornej ludzkiej kontroli nad ASD, jak w wypadku polskiego systemu do profilowania bezrobotnych. Czynniki te ostatecznie mogą utrudniać ustalenie wpływu omawianych systemów na końcowy efekt marginalizacji i wykluczenia społecznego, a także upośledzać skuteczność i realność testu proporcjonalności.

Przyjmując, mimo przedstawionych teoretycznych i praktycznych utrudnień oraz ułomności europejskiego prawa antydyskryminacyjnego, zasadniczą dopuszczalność realizacji ochrony przed dyskryminującymi ASD, należy zauważyć, że również niewyczerpujący katalog niedozwolonych cech chronionych stwarza kolejne problemy ze stosowaniem omawianych regulacji. Mianowicie krytycznie trzeba ocenić brak jednolitych wytycznych co do techniki i metody konstruowania przez sądy lub prawodawstwo dodatkowych, niewymienionych w

[291] *Handbook on European Nondiscrimination Law*, European Union Agency for Fundamental Rights and Council of Europe, European Union and Council of Europe, Brussels 2018, s. 45.
[292] Ch. Tobler, *Limits and Potential of the Concept of Indirect Discrimination*, European Commission, Directorate-General for Employment, Social Affairs and Inclusion 2008, s. 35, https://data.europa.eu/doi/10.2767/56607 (dostęp: 4.12.2021).

88:67315

cytowanych przepisach, kryteriów. Jak już wskazywałam, ASD nie będą różnicowały, opierając się na jednej kategorii chronionej, a na przenikających się zespołach informacji zaobserwowanych w dużych zbiorach danych, częstokroć wykraczających poza aktualne rozumienie cechy chronionej. Z tego względu problem niejednolitego podejścia sądów do możliwości i sposobu uzupełniania katalogów cech chronionych o nowe kryteria jest niezwykle istotny[293].

Analizując orzecznictwo europejskie, można zaobserwować powściągliwość TSUE w poszerzaniu katalogu cech chronionych. Przykładowo Trybunał, odmawiając uznania otyłości za cechę chronioną, w wyroku z 2014 r. wskazał, że należy poszukiwać łącznika otyłości z ewentualną niepełnosprawnością[294]. Również w 2014 r. TSUE nie uznał macierzyństwa za osobne zakazane kryterium różnicujące, odwołując się do różnicowania kobiet w okresie macierzyństwa ze względu na płeć[295]. Trybunał orzekł, że:

> art. 13 WE, obecnie art. 19 TFUE, który zawiera wyłącznie uregulowanie kompetencji Wspólnoty i na podstawie którego przyjęte zostały dyrektywy 2000/43 i 2000/78, również nie odnosi się do dyskryminacji ze względu na przynależność do danej kategorii społecznozawodowej lub miejsca pracy, a zatem ani wspomniany art. 13 WE, ani art. 19 TFUE nie mogą nawet stanowić podstawy prawnej dla przepisów Rady mających na celu zwalczanie tego typu dyskryminacji[296].

Analizując orzecznictwo, można się też spotkać z bardziej aktywnym realizowaniem postulatów równościowych. W wyroku *Chassagnou i in. v. Francja*[297] Europejski Trybunał Praw Człowieka zakazał dyskryminacji ze względu na wielkość majątku ziemskiego, wskazując, że drobni właściciele gruntów byli zobowiązani do przekazania swoich praw do polowania stowarzyszeniom łowieckim, podczas gdy właściciele dużych majątków mieli prawo do zupełnie swobodnego korzystania ze swoich gruntów, przez co za inną przyczynę dyskryminacji uznano sytuację finansową lub materialną. Z kolei w sprawie *Coleman v. Attridge Law i Steve Law*[298] TSUE przyjął, że zakaz dyskryminacji bezpośredniej przewidziany w dyrektywie Rady 2000/78/WE z 27.11.2000 r., ustanawiającej ogólne warunki ramowe równego traktowania w zakresie zatrudnienia i pracy, nie ogranicza się tylko do osób

[293] J. Maliszewska-Nienartowicz, *Karta Praw Podstawowych*…, preprint 21, par. 26.
[294] Wyrok TSUE z 18.12.2014 r. w sprawie *Fag og Arbejde (FOA), działający w imieniu Karstena Kaltofta, v. Kommunernes Landsforening (KL), działającemu w imieniu Billund Kommune*, skarga nr C-354/13.
[295] Wyrok TSUE z 18.03.2014 r. w sprawie *Z. v. A Government department, The Board of management of a community school*, skarga nr C-363/12.
[296] Wyrok TSUE z 7.07.2011 r. w sprawie *Ministerul Justiţiei şi Libertăţilor Cetăţeneşti v. Ştefan Agafiţei i in.*, skarga nr C-310/10, pkt 35.
[297] Wyrok ETPCz z 29.04.1999 r. w sprawie *Chassagnou i in. v. Francja*, skargi nr 25088/94, 28331/95 i 28443/95.
[298] Wyrok TSUE z 17.07.2008 r. w sprawie *S. Coleman v. Attridge Law i Steve Law*, skarga nr C-303/06.

89:63990

niepełnosprawnych, lecz obejmuje także pracownicę, która sama nie jest niepełnosprawna, lecz jest traktowana przez pracodawcę mniej przychylnie – ze względu na niepełnosprawność jej dziecka – niż inny pracownik w porównywalnej sytuacji[299]. TSUE uznał też, że zakaz dyskryminacji odnosi się do osób, które mimo iż same nie są osobami o określonym pochodzeniu rasowym lub nie należą do danej grupy etnicznej, są jednak poddane mniej przychylnemu traktowaniu bądź znajdują się w szczególnie niekorzystnej sytuacji z jednej z tych przyczyn[300].

Tego rodzaju błędna kwalifikacja jest istotnym problemem również w ramach ASD. Mieszkaniec określonej dzielnicy miasta, uznanej za należącą do mniejszości narodowych, może bowiem zostać błędnie zakwalifikowany przez algorytm jako przedstawiciel tej mniejszości jedynie na podstawie kodu pocztowego. Podobny problem dostrzega art. 14 EKPCz, który obejmuje sytuacje, w których bezprawne traktowanie konkretnej osoby następuje ze względu na status lub chronione cechy osobiste innej jednostki, powiązanej z pokrzywdzoną dyskryminacją[301].

Problem ten jest jednak bardziej złożony. Istnienie cech chronionych, których wykształcenie zawdzięczamy koncepcjom prawnonaturalnym, bardziej zależy od tego, kim jest dana osoba, niż od tego, co uczyniła lub co jej uczyniono[302]. Taka konstrukcja może prowadzić do osłabienia zakazu dyskryminacji w obliczu różnicowania na podstawie innych danych, w tym cech behawioralnych[303]. TSUE wydaje się jednak dostrzegać ten problem, gdyż w ramach analizy charakteru dyskryminacji pośredniej wskazuje on, że zakazane są wszystkie – nawet ukryte – formy dyskryminacji, które przez zastosowanie innych kryteriów różnicowania faktycznie prowadzą do tego samego skutku[304].

Omówione niejednoznaczność i wieloaspektowość kształtu oraz zakresu ochrony prawnej przed dyskryminacją automatyczną jako formą dyskryminacji pośredniej powodują, że sądowa ochrona nie jest efektywnym narzędziem do walki z marginalizacją grup społecznych i korozją społeczną wywoływaną pogłębiającym się zjawiskiem niedozwolonego różnicowania

[299] A. Wróbel, *Dyskryminacja bezpośrednia ze względu na więź?*, „EPS" 9, 2008, s. 1.
[300] Wyrok TSUE z 16.07.2015 r. w sprawie *CEZ Razpredelenie Bylgarija AD v. Komisija za zasztita ot diskriminacija, przy udziale: Anelii Nikołowej, Dyrżawna komisija za energijno i wodno regulirane*, skarga nr C-83/14.
[301] M.A. Nowicki, *Komentarz do Konwencji o ochronie praw człowieka i podstawowych wolności*, [w:] *Wokół Konwencji Europejskiej. Komentarz do Europejskiej Konwencji Praw Człowieka*, Warszawa 2021, preprint 14.
[302] A. Wróbel, *Konwencja o ochronie praw człowieka i podstawowych wolności*, t. 1. *Komentarz do artykułów 1–18*, red. L. Garlicki, Warszawa 2010, preprint 14, nb. 42.
[303] P.T. Kim, *Data-Driven Discrimination at Work*, „William & Mary Law Review" 58, 2017, nr 3.
[304] Wyrok TSUE z 12.02.1974 r. w sprawie *Giovanni Maria Sotgiu v. Deutsche Bundespost*, skarga nr C-152/73.

90:49855

automatycznego. Istotnie większa część bezprawnych działań umyka kontroli społecznej i nie spotyka się z odpowiedzią prawa. Nierzadko jednostki dotknięte dyskryminacją z powodu swojego trudnego położenia finansowego lub społecznego, a także wymogów instytucjonalnych związanych z formułą, kosztem i czasem trwania postępowań, ale również z powodu wykluczenia prawnego czy braku wymiernych korzyści z dochodzenia naruszeń, a nadto z obawy przed intensyfikacją sporu ze stroną silniejszą (np. w relacjach państwo–jednostka, pracodawca–pracownik) nigdy nie zdecydują się na realizację ochrony sądowej. Niestety wskazane obawy nie stanowią jedynie teoretycznych rozważań. Seria spraw *Defrenne*[305], która trwała ponad dziesięć lat i w której sąd trzykrotnie zwracał się do TSUE o wydanie orzeczenia w sprawie różnych aspektów zasady równości wynagrodzeń, jest dowodem na uciążliwość procedur i ich nieadekwatność do masowości zagrożenia związanego z automatycznym podejmowaniem decyzji[306].

Na taki wniosek wpływ ma również sama formuła gwarancji prawnych statuowanych Kartą Praw Podstawowych. Zasada niedyskryminacji wiąże bowiem państwa członkowskie tylko wtedy, gdy rozpatrywana sytuacja krajowa wchodzi w zakres stosowania prawa unijnego. Stwierdzenie, że w sprawie znajduje zastosowanie ogólna zasada prawa, nie przesądza o tym, że może być ona uznana za normatywną podstawę rozstrzygnięcia w krajowym sporze[307]. Realnym utrudnieniem w upowszechnieniu europejskich kierunków orzeczniczych niedyskryminacji jest przy tym konieczność zbadania i ustalenia związku dyskryminacji z innym przepisem prawa UE tylko po to, aby otworzyć drogę do zastosowania unijnych praw podstawowych[308]. Jeżeli TSUE uznałby, że sprawa ma charakter krajowy, to zastosowanie przepisów KPP byłoby wyłączone, a w efekcie niedopuszczalne byłoby dokonywanie wykładni przepisów przez Trybunał[309].

Ochrona przed niedozwolonym różnicowaniem ASD doznaje zatem daleko idących praktycznych ograniczeń z perspektywy zakresu zastosowania poszczególnych regulacji. Zakaz dyskryminacji nie znajdzie zastosowania, jeśli główna część skargi pokrzywdzonego nie zmieści się w zakresie prawa lub wolności chronionej na podstawie konkretnego aktu

[305] Sprawy *Gabrielle Defrenne vs Société anonyme belge de navigation aérienne Sabena* zakończone wyrokami TSUE z: 8.04.1976 r., sygn. Defrenne (43/75) oraz 15.06.1978 r., sygn. Defrenne (149/77).
[306] R. Xenidis, L. Senden, *op. cit*.
[307] R. Grzeszczak, A. Szmigielski, *Sądowe stosowanie Karty Praw Podstawowych UE w odniesieniu do państw członkowskich – refleksje na podstawie orzecznictwa Trybunału Sprawiedliwości i praktyki sądów krajowych*, „EPS" 10, 2015, s. 13.
[308] *Ibidem*.
[309] *Ibidem*.

91:10029

prawnego. Należy ponadto zaznaczyć, że mimo horyzontalnego skutku zakazu dyskryminacji uznanego w sprawie *Mangold v. Helm* oraz wyjaśnienia, że art. 21 Karty ma bezpośredni skutek horyzontalny w sprawie *Egenberger v. Evangelisches Werk für Diakonie und Entwicklung*, Trybunał Sprawiedliwości wyjaśnił wcześniej (w sprawach *Chacón Navas v. Eurest Colectividades*, *Ana de Diego Porras v. Ministerio de Defensa*[310] czy *Coleman v. Attridge Law i Steve Law*[311]), że tylko te przyczyny, które znajdują wyraz w prawie wtórnym, mogą być uznane za chronione przez dyrektywy[312]. Poszerzanie zakresu pojęcia dyskryminacji na drodze wykładni sądowej wydaje się niewystarczające.

Aktualnie dyskryminacja ze względu na rasę i pochodzenie etniczne jest zakazana w odniesieniu do zatrudnienia, pracy, szkoleń zawodowych oraz w obszarach z nimi związanych, takich jak ochrona socjalna, opieka zdrowotna, edukacja oraz dostęp do towarów i usług, w tym mieszkań[313]. Dyskryminacja ze względu na płeć jest zakazana w kontekście tych samych obszarów, z wyłączeniem edukacji, mediów, reklamy[314]. Z kolei – jak wskazuje Monika Domańska – dyskryminacja z uwagi na religię i światopogląd, wiek, niepełnosprawność czy orientację seksualną jest zakazana jedynie w zatrudnieniu, pracy i szkoleniach zawodowych[315]. W dużym stopniu zakres stosowania poszczególnych źródeł prawa rozstrzyga zatem o granicach ochrony przed dyskryminacją, przenosząc dyskusję o ujęciu formy dyskryminacji automatycznej na dalszą perspektywę.

Tego typu ograniczenia dostrzegane są nie tylko na gruncie przepisów Unii Europejskiej, ale również Europejskiej konwencji praw człowieka, aczkolwiek regulacja ta przewiduje możliwość rozszerzenia stosowania przepisu na inne dodatkowe prawa, tylko jeśli wchodzą

[310] W sprawie *Ana de Diego Porras v. Ministerio de Defensa* (skarga nr C-596/14) Trybunał uznał, że odmowa przedłużenia umowy pracowniczej ze względu na ciążę pracownicy jest naruszeniem prawa Unii Europejskiej.
[311] Sprawa ta dotyczyła dyskryminacji przez asocjację, a konkretnie – interpretacji dyrektywy UE dotyczącej równego traktowania w zatrudnieniu. We wskazanej sprawie Sharon Coleman pozwała pracodawcę ze względu na dyskryminację z powodu niepełnosprawności swojego syna. W wyroku Trybunał uznał, że żadne postanowienie traktatu nie zabrania dyskryminacji ze względu na chorobę jako taką. Jeśli chodzi o ogólną zasadę niedyskryminacji, zasada ta obowiązuje państwa członkowskie, w których sytuacja krajowa będąca przedmiotem postępowania głównego wchodzi w zakres prawa wspólnotowego. Nie wynika jednak z tego, że zakres dyrektywy 2000/78 powinien zostać rozszerzony w drodze analogii poza dyskryminację ze względu na przyczyny wymienione wyczerpująco w jej art. 1.
[312] J. Gerards, R. Xenidis, *op. cit.*; zob. wyrok TSUE z 22.11.2005 r. w sprawie *Werner Mangold v. Rüdiger Helm*, skarga nr C-144/04; wyrok TSUE z 17.04.2018 r. w sprawie *Vera Egenberger v. Evangelisches Werk für Diakonie und Entwicklung eV*, skarga nr C-414/16; wyrok TSUE z 11.07.2006 r. w sprawie *Sonia Chacón Navas v. Eurest Colectividades SA*, skarga nr C-13/05; wyrok TSUE z 18.12.2014 r. w sprawie *Fag og Arbejde (FOA) v. Kommunernes Landsforening (KL)*, skarga nr C-354/13; wyrok TSUE z 14.09.2016 r. w sprawie *Ana de Diego Porras v. Ministerio de Defensa*, skarga nr C-596/14; wyrok TSUE z 17.07.2008 r. w sprawie *S. Coleman v. Attridge Law*…
[313] M. Domańska, *op. cit.*, rozdz. 1.4.1.
[314] *Ibidem*.
[315] R. Grzeszczak, A. Szmigielski, *op. cit.*, s. 13.

92:41954

one w ogólny zakres danego przepisu Konwencji. Jeżeli zatem główna część skargi nie mieści się w zakresie prawa lub wolności chronionej na podstawie Konwencji, to art. 14 EKPCz nie może być stosowany[316]. Z tej przyczyny art. 14 jest aplikowany przeważnie w związku z innymi przepisami Konwencji i kwalifikowany jako przepis akcesoryjny[317].

Wymóg ten co prawda ulega obecnie złagodzeniu, aczkolwiek nie można jednoznacznie określić, czy i jak długo aktualne orzecznictwo Trybunału będzie utrzymywało taki kierunek. Należy jednak zaznaczyć, że gwarancje praw i wolności muszą zawsze uwzględniać obowiązek ochrony przed dyskryminacją. Gwarancje wyznaczone przez art. 14 EKPCz są bowiem inherentną częścią wszystkich praw i wolności niezależnie od ich natury[318]. Skutkiem tego jest ograniczenie stosowania art. 14 Konwencji w obszarach zatrudnienia czy praw socjalnych wyłączonych z zakresu przedmiotowego Konwencji[319].

Kolejnym ważnym problemem jest ograniczona możliwość powoływania się wprost na unijne prawo równościowe przez osoby fizyczne. Wiele przepisów antydyskryminacyjnych w zakresie zakazu różnicowania ze względu na płeć, wiek czy rasę zostało bowiem umiejscowione na poziomie dyrektyw, które same w sobie nie mają bezpośredniego skutku horyzontalnego i zasadniczo będą mogły mieć zastosowanie do stosunków prywatnych dopiero po ich transpozycji do prawa krajowego. W orzecznictwie[320] utrwalono, że ogólna zasada niedyskryminacji i art. 21 Karty Praw Podstawowych Unii Europejskiej mają horyzontalny skutek bezpośredni, jednak nie rozstrzyga to kontrowersji wobec legitymacji w sporach osób fizycznych (np. konsumentów) z podmiotami gospodarczymi z sektora prywatnego, które stosują automatyczne systemy decyzyjne w swoich przedsiębiorstwach. Koniecznym uzupełnieniem kontroli sądowej powinno być zatem stworzenie ram i zaangażowanie instytucjonalnych narzędzi związanych z ogólnym monitorowaniem przestrzegania prawa unijnego, w tym przepisów antydyskryminacyjnych. Orzeczenia Trybunału nie mają bowiem bezpośredniego wpływu na sektor prywatny w tym znaczeniu, że jedynie państwa członkowskie mogą być pociągnięte do odpowiedzialności, gdy przepisy unijne nie są

[316] Wyrok ETPCz z 22.03.2012 r. w sprawie *Konstantin Markin v. Rosja*, skarga nr 30078/06, pkt 124.
[317] Por. wyrok ETPCz z 12.07.2001 r. w sprawie *Prince Hans-Adam II of Liechtenstein v. Niemcy* (skarga nr 42527/98, par. 91–94), w którym Trybunał stwierdził, że brak związku między prawami i wolnościami gwarantowanymi przez Konwencję a dyskryminacją powoduje, że art. 14 nie ma zastosowania.
[318] M. Domańska, *op. cit.*, rozdz. 4.2.
[319] A. Wróbel, *Konwencja o ochronie praw człowieka*…, t. 1, preprint 14, nb. 17.
[320] Wyrok TSUE z 22.11.2005 r. w sprawie *Werner Mangold*…; wyrok TSUE z 19.01.2010 r. w sprawie *Seda Kücükdeveci v. Swedex GmbH & Co. KG*, skarga nr C-555/07; wyrok TSUE z 19.04.2016 r. w sprawie *Dansk Industri (DI), działający w imieniu Ajos A/S v. Estate of Karsten Eigil Rasmussen*, skarga nr C-441/14; wyrok TSUE z 17.04.2018 r. w sprawie *Vera Egenberger v. Evangelisches Werk für Diakonie und Entwicklung e.V.*, skarga nr C-414/16.

93:31168

realizowane lub transponowane (tak art. 258 TFUE). Na państwa członkowskie powinny być więc nałożone obowiązki kontrolne i nadzorcze w ramach egzekwowania od podmiotów prywatnych, w tym korzystających z ASD, by przestrzegały unijnych gwarancji o niedyskryminacji[321]. O ile jednak obowiązki związane z równościowymi standardami obrotu gospodarczego nie zostaną określone na podstawie samego prawa UE w sposób egzekwowalny i weryfikowalny, o tyle nawet realizacja nadzoru będzie miała charakter pozorny. Podobne ograniczenia dostrzegane są na gruncie regulacji EKPCz[322].

Z prawnego punktu widzenia dyskryminacja w ASD odnosi się zatem tylko do nieuzasadnionego niekorzystnego traktowania lub niekorzystnej sytuacji doświadczanej przez określone kategorie ludzi o cechach wprost wskazanych jako chronione lub takich, które ze względu na otwarty katalog zostaną w toku stosowania prawa zrównane z cechami chronionymi.

Na podstawie zaprezentowanej do tej pory analizy dostrzegam, że europejskie prawo antydyskryminacyjne może mieć problemy z wykryciem i realizacją ochrony przed różnicowaniem algorytmicznym ze względu na masowość decyzji, epistemiczną nieprzejrzystość procesów algorytmicznych, nowe niejawne formy różnicowania i nieadekwatność narzędzi regulacyjnych, a także brak efektywnych instrumentów nadzoru. Pozytywnie oceniając wkład orzecznictwa Trybunałów w adaptację regulacji do zmieniających się stosunków społeczno-gospodarczych, widzę ułomności rozwiązania opartego na ograniczeniu ochrony przez uzależnienie jej zakresu od wszczęcia i prowadzenia sporów przed Trybunałami. Aktywny udział orzecznictwa sądowego jest z pewnością niezbędny, ale nie może być on jedynym gwarantem realizacji postulatów równościowych. Należy wobec tego przeanalizować inne szczególne regulacje prawne mogące uzupełniać zakaz dyskryminacji o szczegółowe obowiązki i prawa związane z przetwarzaniem dużych ilości danych w systemach automatycznych.

3. Ochrona danych osobowych

Jednym z podstawowych fragmentów regulacji odpowiedzialnej za zabezpieczenie gwarancji obywatelskich w związku z przetwarzaniem i analizą danych jest prawo ochrony danych

[321] R. Xenidis, L. Senden, *op. cit*.
[322] M.A. Nowicki, *op. cit.*, preprint 21.

94:87881

osobowych uregulowane w rozporządzeniu Parlamentu Europejskiego i Rady (UE) 2016/679 z 27.04.2016 r. w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46/WE[323]. Ochrona danych osobowych w jej obecnym europejskim kształcie może się przyczyniać do ochrony jednostki przed dyskryminacją nie tylko przedmiotowo, ale również poprzez dostarczanie innowacyjnych narzędzi regulacyjnych. Komisja Europejska wskazała nawet ten obszar jako jeden z kluczowych fragmentów prawa w zakresie budowania europejskiej strategii cyfryzacji, w tym ram sztucznej inteligencji.

Analizując zakres tej regulacji, należy w pierwszej kolejności wskazać na podobieństwa przedmiotowe i wspólny mianownik przetwarzania danych w ASD i poza nimi. RODO zakłada bowiem, że każdy proces przetwarzania danych osobowych może prowadzić do ryzyka naruszenia praw i wolności podmiotów danych, a jednocześnie charakter tego ryzyka uniemożliwia statyczne zarządzanie nim za pomocą sztywno narzuconych regulacji[324]. Podobne wnioski można wyciągnąć na gruncie zagrożeń związanych z analizą danych przez ASD, szczególnie te opierające się na algorytmach czarnoskrzynkowych.

Mechanizmy regulacyjne wprowadzone RODO powinny zatem odpowiadać na potrzeby związane z immanentnym, choć równocześnie niedostrzegalnym dla ludzi ryzykiem naruszenia praw i wolności. Strategia regulacyjna umożliwiająca proaktywny i elastyczny dobór narzędzi dla uczestników obrotu do realizacji celów prawodawcy poprzez nacisk na wewnętrzną samoregulację jest przy tym koniecznym i właściwym kierunkiem również z perspektywy faktycznych ograniczeń ludzkich i niemożliwości realnego nadzorowania procesów różnicowania w automatycznych systemach decyzyjnych[325]. Na konieczność redefiniowania strategii regulacyjnej ma także wpływ niejawność i masowość procesów, co dowodzi podobieństwa kategoryzowania dokonywanego na gruncie przepisów RODO do różnicowania automatycznego. Opisane wcześniej ryzyka związane z nieprawidłowościami w toku różnicowania wymuszają przy tym zniesienie formalizmu i państwowej centralizacji nadzoru, jako że ograniczenia po stronie regulatora uniemożliwiają operowanie narzędziami tradycyjnie rozumianej władzy publicznej. Mimo tego słusznego kierunku RODO zawiera

[323] Dz.Urz. UE L 119.
[324] M. Pichlak, K. Gaczoł, *Refleksyjna regulacja ryzyka w ochronie danych osobowych. podstawy teoretyczne i praktyka działalności prezesa UODO*, Łódź 2022, s. 88.
[325] *Ibidem*.

95:98670

istotne ograniczenia, które wpływają na upośledzenie stosowania mechanizmów ochronnych w kluczowych dla dyskryminacji automatycznej aspektach.

Ochrona danych osobowych dotyczy informacji o zidentyfikowanej lub możliwej do zidentyfikowania osobie fizycznej[326], automatyczne systemy decyzyjne zaś opierają się nie tylko na danych osobowych, ale też na uczeniu się poprzez analizę dużych zbiorów danych o różnym charakterze i pochodzeniu. Zakwalifikowanie danych ujętych w *big data* do kategorii prawnej danych osobowych może implikować stosowanie istniejących w RODO ograniczeń prawnych, a wobec tego stanowić oczekiwaną odpowiedź prawa na zaobserwowany problem automatycznej dyskryminacji; w pozostałym zakresie RODO nie znajdzie jednak zastosowania do ochrony praw i wolności jednostki. Tym samym jedynie gdy system automatyczny wykorzystuje chociaż w części lub na którymś etapie dane osobowe – czy to na poziomie jego szkolenia, testowania, czy wdrożenia – do całego procesu znajdą zastosowanie narzędzia wprowadzone przez RODO. W przypadku analizy innych informacji[327], w tym danych niepozwalających nawet na pośrednią identyfikację konkretnej osoby fizycznej, system z uwagi na wykorzystywanie otwartej kategorii danych nie będzie podlegał regulacji RODO[328]. Niestety przeciętny użytkownik systemu automatycznego, a nawet jego operator nie będzie w stanie ustalić rodzaju danych ani ich pochodzenia.

Trzeba też brać pod uwagę, że w przypadku danych zbieranych internetowo można wyróżnić ich trzy podstawowe źródła. Po pierwsze, będą to dane wprowadzone przez użytkowników świadomie (np. blogi, formularze kontaktowe, konta społecznościowe); po drugie – dane zaobserwowane (np. dane behawioralne, geolokalizacja użytkownika, historia przeglądarki, czas trwania aktywności); natomiast trzeci rodzaj danych to dane wyinterpretowane (np. dane wywnioskowane przez algorytmy, dotyczące statusu rodzinnego, liczby dzieci czy stanu majątkowego). Systemy automatyczne mogą bowiem analizować również nieustrukturyzowane informacje, takie jak metadane generowane z Internetu rzeczy[329].

[326] Art. 4 pkt 1 RODO.

[327] Dane nieosobowe to np. dane statystyczne czy zanonimizowane dane osobowe. W stosunku do danych nieosobowych zastosowanie znajduje m.in. rozporządzenie Parlamentu Europejskiego i Rady (UE) 2018/1807 z 14.11.2018 r. w sprawie ram swobodnego przepływu danych nieosobowych w Unii Europejskiej, Dz.Urz. L 303 z 28.11.2018 r., s. 59–68.

[328] M. Porzeżyński, *Problematyka własności danych generowanych w ramach Internetu rzeczy*, „Państwo i Prawo" 2, 2021, s. 123–137.

[329] Internet rzeczy (ang. *Internet of things* – IoT) to sieć łącząca urządzenia (rzeczy) charakteryzujące się autonomicznym (niewymagającym zaangażowania człowieka) działaniem w zakresie pozyskiwania, udostępniania, przetwarzania danych lub wchodzenia w interakcje z otoczeniem pod wpływem tych danych, czyli *de facto* wykorzystującym sztuczną inteligencję lub uczenie maszynowe. W warstwie architektonicznej IoT jest traktowane jako architektura architektoniczna, oparta na czterech warstwach: 1. sprzęt, 2. komunikacja, 3.

96:84803

Dodatkowym źródłem danych jest ich automatyczne przypisywanie użytkownikom na skutek procesów profilowania.

Musimy ponadto pamiętać, że wciąż brakuje jednolitego podejścia do zagadnienia, czy profile algorytmiczne lub algorytmiczne wnioskowanie na temat osoby fizycznej są danymi osobowymi, czy też nie[330]. W literaturze prezentowane jest także stanowisko, które dopuszcza objęcie regulacją wszelkich danych, w tym profili, które potencjalnie wpływają na innych ludzi[331].

Podczas gdy ogólne przetwarzanie danych ma na celu umożliwienie wnioskowania o pewnych stanach, celem przetwarzania danych osobowych w ASD jest kategoryzowanie ludzi i tworzenie profili zarówno indywidualnych, jak i grupowych[332]. Mają one obejmować cechy istotne dla procesu decyzyjnego, umożliwiające racjonalizację środków i maksymalizację efektywności procesów. Dlatego też RODO może być nie tylko fragmentem odpowiedniej regulacji, ale przede wszystkim wzorem realizacji nowej filozofii strategii regulacyjnej. Przypisywanie ważnych z punktu widzenia procesu decyzyjnego atrybutów konkretnym podmiotom danych opiera się przy tym na zestawianiu profili indywidualnych z grupowymi[333]. Zindywidualizowany profil w zestawieniu z profilem grupowym pozwala na przypisanie użytkownikowi nieujawnionych pierwotnie cech – jeżeli dana osoba zgodnie z analizą ma cechy A i B, to powinna dysponować także cechami C i D wynikającymi z profilu grupowego. Tym samym tworzone są nowe dane o osobie fizycznej, które nie zostały przekazane przez nią bezpośrednio[334].

Proces ten zbliża przetwarzanie danych osobowych do predykcji stosowanej przez automatyczne systemy decyzyjne. Podmiot danych nie ma przy tym wpływu na czynniki, które kształtują jego profil, a wręcz często w ogóle ich nie zna, nie ma też wiedzy o relacjach

oprogramowanie oraz 4. integracja; por. M. Nowakowski, *FINTECH – technologia, finanse, regulacje. Praktyczny przewodnik dla sektora innowacji finansowych*, Warszawa 2020, rozdz. X.

[330] M. Mann, T. Matzner, *Challenging Algorithmic Profiling: The Limits of Data Protection and Anti-Discrimination in Responding to Emergent Discrimination*, „Big Data & Society" 6, 2019, nr 2, s. 1–11.

[331] N. Purtova, *The Law of Everything: Broad Concept of Personal Data and Future of EU Data Protection Law*, „Law, Innovation and Technology" 10, 2018, nr 1, s. 40–81.

[332] Profil to zestaw danych charakteryzujący kategorie osób, który ma zostać zastosowany w odniesieniu do danej osoby; por. rekomendacja CM/Rec (2010) 13 Komitetu Ministrów państw członkowskich w sprawie ochrony osób w związku z automatycznym przetwarzaniem danych osobowych podczas tworzenia profili, przyjęta przez Komitet Ministrów 23.11.2010 r. podczas 1099. posiedzenia Wiceministrów.

[333] M. Ciechomska, *Prawne aspekty profilowania oraz podejmowania zautomatyzowanych decyzji w ogólnym rozporządzeniu o ochronie danych osobowych*, „EPS" 5, 2017, s. 37–42.

[334] Tworzenie profili oznacza automatyczną technikę przetwarzania danych, polegającą na przypisaniu danej osobie „profilu" w celu podejmowania dotyczących jej decyzji bądź analizy lub przewidywania jej preferencji, zachowań i postaw; por. rekomendacja CM/Rec (2010).

97:73853

pomiędzy tymi czynnikami[335]. Użytkownik może jednocześnie zostać błędnie lub nieprecyzyjnie skategoryzowany, nie będąc przedstawicielem rozpoznawalnej grupy społecznej. Taka kategoryzacja zaś może mieć bezpośredni wpływ na jego prawa i obowiązki.

Ponadto powszechna dostępność danych w Internecie, gromadzonych zarówno wprost od użytkowników, jak i za pośrednictwem różnych urządzeń i narzędzi, oraz zdolność ASD do analizy danych, wyszukiwania korelacji i tworzenia powiązań pozwalają na skuteczną predykcję zachowań obywateli. Ustalony na podstawie analizy danych profil może zatem prowadzić do zamykania ludzi w kategoriach zakładających pewne cechy lub preferencje, czego efektem może być nie tylko ograniczanie wolności wyboru osób fizycznych, ale również segmentowanie dostępu do dóbr i informacji, ale przede wszystkim utrwalanie stereotypów i pogłębianie segregacji społecznej w związku z niemożliwością wyjścia z grupy, do której zostało się przypisanym na skutek przeprowadzonego procesu różnicowania. W efekcie osoby mieszkające w określonych dzielnicach, korzystające z konkretnych świadczeń społecznych lub mające daną historię zatrudnienia czy zdolność kredytową mogą nie mieć dostępu do ofert sprofilowanych pod kątem innych grup społecznych (np. kobiety mogą nie widzieć ofert pracy sprofilowanych pod mężczyzn).

Elementem profilowania może być przy tym prognozowanie, które zwiększa ryzyko nieprawidłowości. W literaturze podjęto nawet próbę stypizowania zagrożeń związanych z profilowaniem, po której wskazano na:

1) nieprzejrzystość algorytmów, na których oparte jest tworzenie profili oraz budowanie korelacji między stworzonym profilem i np. oferowaną danej osobie usługą;
2) potencjalne naruszenie prywatności poprzez zbieranie danych bez wiedzy osób, których dane dotyczą;
3) niezrozumienie tego, co wynika z profilowania;
4) utrwalanie istniejących stereotypów i segregację społeczną;
5) możliwość trwałego powiązania danej osoby z określoną kategorią i przypisania jej do konkretnych preferencji, co może być źródłem dyskryminacji[336].

[335] D. Lubasz, Big Brother Is Watching You. *Profilowanie i zautomatyzowane podejmowanie decyzji w kontekście zasad legalności i przejrzystości*, [w:] *Rok RODO 2019*, red. W. Wiewiórowski, H. Wolska, Warszawa 2019, s. 115–134.

[336] J. Niklas, *Profilowanie w kontekście ochrony danych osobowych i zakazu dyskryminacji*, Polskie Towarzystwo Prawa Antydyskryminacyjnego [b.r.w.], http://ptpa.org.pl/site/assets/files/publikacje/opinie/Opinia_profilowanie_w_kontekscie_ochrony_danych_osobowych_i_zakazu_dyskryminacji.pdf (dostęp: 4.12.2021).

98:96880

W tym kontekście można uznać, że współczesna inwigilacja, polegająca na przetwarzaniu na masową skalę danych pozostawianych m.in. w Internecie przez użytkowników, wzmacnia istniejący i wyznacza nowy krajobraz segregacji społecznej, o czym szerzej pisał David Lyon[337]. Segregacja i wykluczenie są zaś nierozerwalnie związane z dyskryminacją. Z tego względu jednym z założeń RODO jako narzędzia do ograniczania nieprawidłowości związanych z przetwarzaniem danych było zmniejszenie niebezpieczeństwa naruszenia prawa do prywatności i przeciwdziałanie procesom dyskryminacji. Regulacja ta posługuje się jednak tymi pojęciami niejako mechanicznie. Brak refleksji nad adekwatnością istniejącego pojęcia dyskryminacji do procesów różnicowania algorytmicznego czyni aktualnymi zarzuty sformułowane we wcześniejszych rozdziałach również na gruncie stosowania RODO. O ile bowiem różnicowanie algorytmiczne będzie się wymykać definicji dyskryminacji, o tyle użytkownicy systemów nie będą mogli przeciwdziałać niesprawiedliwemu różnicowaniu, a w konsekwencji będą narażeni na postępującą marginalizację i wykluczenie.

RODO posługuje się pojęciem automatycznych decyzji w kontekście regulacji profilowania danych osobowych[338], przez co rozumie wykorzystanie danych osobowych do oceny niektórych czynników osobowych osoby fizycznej, szczególnie do analizy lub prognozy aspektów dotyczących efektów pracy tej osoby fizycznej, jej sytuacji ekonomicznej, zdrowia, osobistych preferencji, zainteresowań, wiarygodności, zachowania, lokalizacji lub ścieżek przemieszczania się[339]. Chodzi zatem o analizę danych z wykorzystaniem matematycznych i statystycznych modeli, wykonywaną przez lub na polecenie administratora danych[340], polegającą na zbieraniu niepowiązanych z sobą danych dotyczących tej samej osoby lub na wnioskowaniu o nieznanych cechach danej osoby na podstawie cech ustalonych wcześniej[341].

Celem profilowania jest przewidywanie zachowań osób fizycznych[342]. W praktyce ma ono bardzo szerokie zastosowanie – wykorzystuje się je do tworzenia lepiej adresowanych reklam

[337] Z. Bauman, D. Lyon, *Płynna inwigilacja. Rozmowy*, przeł. T. Kunz, Kraków 2013, s. 26.
[338] RODO posługuje się pojęciem profilowania w motywach 71 i 72, a jego definicja legalna została zawarta w art. 4 pkt 4.
[339] Art. 4 pkt 4 RODO.
[340] Administrator to osoba fizyczna lub prawna, organ publiczny, jednostka lub inny podmiot, który samodzielnie lub wspólnie z innymi ustala cele i sposoby przetwarzania danych osobowych. Jeżeli cele i sposoby takiego przetwarzania są określone w prawie Unii Europejskiej lub państwa członkowskiego, to również w prawie UE lub państwa członkowskiego może zostać wyznaczony administrator lub mogą zostać określone konkretne kryteria jego wyznaczania; zob. art. 4 ust. 7 RODO.
[341] W. Wiewiórowski, *Profilowanie osób na podstawie ogólnodostępnych danych*, [w:] *Prywatność a ekonomia. Ochrona danych osobowych w obrocie gospodarczym*, red. A. Mednis, Warszawa 2013, s. 26.
[342] Procesy te mają wiele zastosowań komercyjnych, przykładowo mogą być wykorzystywane do lepszego dzielenia rynków na segmenty oraz dostosowywania ofert i produktów do indywidualnych potrzeb. Mogą także przynosić korzyści w medycynie, edukacji i transporcie – tak: Grupa Robocza Art. 29 rozporządzenia Parlamentu

99:35120

w sektorze finansowo-bankowym (np. profile buduje się na podstawie analizy danych z kart kredytowych czy danych z rejestrów zadłużeń), lepiej dobranych instrumentów ubezpieczeniowych w sektorze ubezpieczeniowym (profilowanie obejmuje dane dotyczące ryzyka ubezpieczeniowego), lepiej dopasowanej oferty w sektorze handlowym lub usługowym (profiluje się, opierając się na treści przeglądanych stron internetowych; ocenia się przyzwyczajenia zakupowe, popularność poszczególnych produktów lub lokalizacji z podziałem na użytkowników) czy też lepiej prezentującej się oferty wyborczej (następuje np. profilowanie wyborców pod kątem ich oczekiwań lub formułowanie obietnic wyborczych dostosowanych do preferencji wyborców)[343]. Profilowanie jest wykorzystywane również w sektorze publicznym, co wprost ukazały przypadki SyRI i polskiego systemu do aktywizacji bezrobotnych. Poza obszarem zabezpieczeń społecznych z kolei wykorzystuje się je np. do: przeciwdziałania terroryzmowi poprzez analizę informacji o tym, skąd, gdzie, jak często i z kim dana osoba podróżuje; identyfikowania nadużyć finansowych poprzez monitoring transakcji i budowanie profilu właściciela konta; a także identyfikowania przypadków handlu ludźmi[344].

Profilowanie często jest przy tym nieprawidłowo utożsamiane z automatycznym podejmowaniem decyzji. Tymczasem zautomatyzowane podejmowanie decyzji może, ale nie musi obejmować profilowania, podczas gdy profilowanie może zachodzić bez podejmowania zautomatyzowanych decyzji[345]. Oba procesy mogą istnieć rozłącznie. Natomiast regulacja RODO będzie mogła stanowić odpowiedź prawa na problem dyskryminacji automatycznej, o ile profilowanie będzie prowadziło do zautomatyzowanego podejmowania decyzji, oczywiście przy spełnieniu wstępnego wymogu przetwarzania danych osobowych w ramach procesu. RODO przewiduje w takiej sytuacji prawo podmiotu danych do niepodlegania decyzji, która opiera się wyłącznie na zautomatyzowanym przetwarzaniu, w tym profilowaniu, jeżeli decyzja wywołuje wobec tej osoby skutki prawne lub w podobny sposób istotnie na nią wpływa[346]. Oznacza to, że o ile w procesie decyzyjnym brał udział człowiek i wykonywane przez niego czynności miały realny wpływ na decyzję, w tym np. nie sprowadzały się do wprowadzania

Europejskiego i Rady (UE) 2016/679 z 27.04.2016 r. w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46/WE (ogólne rozporządzenie o ochronie danych), wytyczne WP 251 rev. 01.

[343] M. Kibil, *Profilowanie i automatyczne podejmowanie decyzji*, LEX/el. 2018.

[344] *Ibidem*.

[345] P. Barta *et al.*, *Ogólne rozporządzenie o ochronie danych osobowych. Ustawa o ochronie danych osobowych. Wybrane przepisy sektorowe. Komentarz*, red. P. Litwiński, Warszawa 2021, preprint 22, nb. 16.

[346] Art. 22 ust. 1 RODO.

100:10780

danych do systemu lub nadzoru nad nim[347], to podmiot danych nie może żądać zaprzestania przetwarzania jedynie z uwagi na automatyzację procesu decyzyjnego. Wyłączone z ograniczeń będą zatem np. systemy eksperckie lub hybrydowe, które sugerują rozstrzygnięcia, podczas gdy ostateczny wybór drogi postępowania jest pozostawiony człowiekowi.

Kolejnym aspektem wpływającym na stosowanie wskazanej regulacji będą skutki decyzji. Przepisy RODO stosuje się bowiem jedynie wtedy, gdy: 1. zautomatyzowane podejmowanie decyzji wywołuje skutki prawne wobec osoby, której dane dotyczą; lub 2. zautomatyzowane podejmowanie decyzji wpływa istotnie na osobę, której dane dotyczą, w podobny sposób do skutków prawnych[348]. W opinii Grupy Roboczej Art. 29[349] pojęcie skutków prawnych należy przy tym wykładać szeroko jako sytuację ingerencji w prawa osoby fizycznej. W literaturze z kolei postuluje się rozumienie pojęcia skutków prawnych węziej – jako powstanie, zmianę lub ustanie stosunku prawnego[350]. Jeżeli zaś chodzi o skutki podobne do skutków prawnych, to jako przykłady podaje się odmowę zawarcia umowy z podmiotem danych, automatyczne ustalanie wysokości składki ubezpieczeniowej czy sposobu lub wysokości zabezpieczenia kredytu.

W związku z tym w literaturze przedmiotu[351] uznaje się, że prezentowanie reklamy na podstawie ustalonego dzięki przetwarzaniu danych profilu osoby fizycznej nie rodzi skutków prawnych i nie wpływa w podobny istotny sposób na osobę, której dane dotyczą. Niemniej jednak można się spotkać z poglądem podążającym za szerokim rozumieniem skutków podobnych do skutków prawnych, tj. że decyzje oparte na profilowaniu podejmowane w ramach działań marketingowych, które powodują poważny uszczerbek w majątku osoby fizycznej (reklama szybkiej pożyczki na wysoki procent wyświetlona osobie będącej w momencie manii albo trudności finansowych), mogą doprowadzić do pogorszenia stanu psychicznego czy fizycznego osoby fizycznej (reklama ryzykownych środków na

[347] Grupa Robocza Art. 29 przyjęła również, że niedopuszczalne byłyby sytuacje, w których udział człowieka w procesie podejmowania decyzji kreowany byłby sztucznie, np. zatwierdzanie profili przygotowanych w sposób automatyczny, bez wpływu na ich treść; por. wytyczne WP 251 rev. 01, s. 10.
[348] P. Barta *et al.*, *op. cit.*, preprint 22, nb. 7.
[349] Grupa Robocza Art. 29 jako przykład podała przyznanie lub odmowę przyznania świadczenia socjalnego, odmowę zgody na przekroczenie granicy, zastosowanie wzmożonych środków bezpieczeństwa czy automatyczne wyłączenie telefonu na skutek niezapłacenia rachunku w terminie; zob. wytyczne WP 251 rev. 01, s. 10.
[350] W. Chomiczewski, *Profilowanie – wyjaśnienie pojęcia,* [w:] *Polska i europejska reforma ochrony danych osobowych*, red. D. Lubasz, E. Bielak-Jomaa, Warszawa 2016, LEX/el., s. 131.
[351] A. Mednis, *Prawo do prywatności a interes publiczny*, Warszawa-Kraków 2006, s. 177.

101:16936

odchudzanie) albo negatywnie wpłynąć na jej sytuację rodzinną lub zawodową, również podlegają pod szczególny reżim RODO[352].

Przyjmuje się też za opinią Grupy Roboczej Art. 29, że każdy skutek prowadzący do wykluczenia lub dyskryminacji będzie skutkiem istotnym w myśl przepisów RODO[353]. Zgodnie z motywem 71 RODO w przypadku automatycznego podejmowania decyzji i profilowania administrator danych powinien bowiem wprowadzać takie środki, które zapobiegną skutkom w postaci dyskryminacji ze względu na „pochodzenie rasowe lub etniczne, poglądy polityczne, wyznanie lub przekonania, przynależność do związków zawodowych, stan genetyczny lub zdrowotny, orientację seksualną". Katalog cech wskazanych w RODO odbiega zatem od KPP – brakuje zwłaszcza odwołania do płci, niepełnosprawności czy wieku jako cech chronionych prawem antydyskryminacyjnym; zakres ochrony jest więc ograniczony.

Procesy prowadzące do wydawania automatycznych decyzji na skutek profilowania osób fizycznych są co do zasady zabronione. RODO przewiduje tymczasem pewne wyjątki od tej reguły, które jednak nie dotyczą dzieci (tj. osób do 16. roku życia). Przede wszystkim aby administrator mógł stosować automatyczne systemy, musi wykazać, że decyzja podjęta w ten sposób: 1. jest niezbędna do zawarcia lub wykonania umowy między osobą, której dane dotyczą, a administratorem; 2. jest dozwolona prawem Unii lub państwa członkowskiego, któremu podlega administrator i które przewiduje właściwe środki ochrony praw, wolności i prawnie uzasadnionych interesów osoby, której dane dotyczą; lub 3. opiera się na wyraźnej zgodzie osoby, której dane dotyczą[354]. Należy przy tym zaznaczyć, że zgodę musi cechować dobrowolność, konkretność, świadomość i jednoznaczność. Wskazane wyjątki nie dotyczą decyzji opierających się na profilowaniu danych osobowych szczególnych kategorii, chyba że podstawą przetwarzania danych osobowych należących do szczególnych kategorii byłby art. 9 ust. 2 lit. a) (przetwarzanie za wyraźną zgodą osoby, której dane dotyczą) lub lit. g) RODO (przetwarzanie niezbędne ze względów związanych z ważnym interesem publicznym, na podstawie prawa Unii lub państwa członkowskiego).

[352] Grupa Robocza Art. 29 ponownie szerzej rozumie podobny wpływ, wskazując, że nie jest wykluczone, że samo prezentowanie reklamy internetowej opierające się na profilowaniu może zostać uznane za wywieranie istotnego wpływu na osobę w sposób podobny do skutków prawnych; por. wytyczne WP 251 rev. 01, s. 11.
[353] A. Mednis, *Istota profilowania osób i jego wykorzystanie w sektorze publicznym*, „ABI-Expert" 1, 2018, s. 12, http://www.abi-expert.pl/wydania/styczen-marzec-2018/art,1986,istota-profilowania-osob-i-jego-wykorzystanie-w-sektorze-publicznym.html (dostęp: 4.12.2021).
[354] Art. 22 ust. 2 RODO.

102:97163

O ile administrator na podstawie zgody lub przydatności do umowy uzyska uprawnienie do posługiwania się decyzją, które opiera się wyłącznie na zautomatyzowanym przetwarzaniu danych, RODO nakłada na niego szczególne obowiązki w zakresie legalności, rzetelności i przejrzystości procesów przetwarzania danych. Po pierwsze, wymóg dotyczący legalności nakłada na administratora obowiązek wprowadzenia środków zapobiegających arbitralnemu, dyskryminującemu traktowaniu podmiotów danych, a w konsekwencji nakazuje implementację adekwatnych rozwiązań i mechanizmów, aby nie dopuścić do negatywnych efektów profilowania lub podejmowania decyzji z naruszeniem wskazanych zasad[355]. Po drugie, administrator ma obowiązek informowania podmiotu danych o automatycznym podejmowaniu decyzji[356], w tym o profilowaniu, oraz musi przekazać podmiotom danych informacje o zasadach podejmowania decyzji, jednocześnie opisując konsekwencje tego przetwarzania dla użytkownika systemu[357]. Po trzecie wreszcie, administrator, który dokonuje systematycznego i kompleksowego profilowania w celu podejmowania decyzji wywołujących skutki prawne lub podobne, musi przeprowadzić we własnym zakresie procedurę wewnętrzną, nazwaną oceną skutków dla ochrony danych. Dodatkowo ma on obowiązek zagwarantować podmiotowi dostęp do informacji dotyczących danych osobowych wykorzystanych do celów profilowania, w tym informacji, jakie kategorie danych zostały wykorzystane do stworzenia profilu[358], a także zapewnić podmiotowi możliwość sprostowania[359], usunięcia[360] lub ograniczenia przetwarzania[361] danych. Administrator musi również wyraźnie (tzn. odrębnie) poinformować osobę, której dane przetwarza, o prawie do sprzeciwu wobec takiego sposobu podejmowania decyzji[362]. W ramach sprzeciwu podmiot danych musi mieć zagwarantowane prawo żądania zaprzestania profilowania z przyczyn związanych ze szczególną sytuacją tej osoby[363] oraz bezwarunkowe prawo żądania zaprzestania profilowania, jeżeli jest ono realizowane na potrzeby marketingu bezpośredniego[364]. Administrator ma przy tym obowiązek zagwarantować podmiotowi danych możliwość uzyskania ludzkiej interwencji w proces

[355] D. Lubasz, M. Namysłowska, *Zasady dotyczące przetwarzania danych osobowych a sztuczna inteligencja w kontekście europejskim*, [w:] *Sztuczna inteligencja*, blockchain*, cyberbezpieczeństwo*…, Legalis/el.
[356] Art. 13 i 14 RODO.
[357] Art. 13 ust. 2 lub art. 14 ust. 2 RODO.
[358] Art. 15 RODO.
[359] Art. 16 RODO.
[360] Art. 17 RODO.
[361] Art. 18 RODO.
[362] Art. 21 RODO.
[363] Art. 21 ust. 1 RODO.
[364] Art. 21 ust. 2 RODO.

103:84011

decyzyjny oraz prawo do wyrażenia własnego stanowiska i zakwestionowania automatycznej decyzji[365].

Obowiązek informacyjny ma szczególnie doniosłe znaczenie dla detekcji i przeciwdziałania naruszeniom prawa. Ma on bowiem zastosowanie do każdej postaci profilowania, nie tylko wtedy, gdy mamy do czynienia z profilowaniem kwalifikowanym[366]. Administrator powinien go realizować podczas zbierania danych osobowych, a gdy profilowanie opiera się na danych „zewnętrznych", niepozyskanych od osoby profilowanej – z chwilą ich pozyskania bądź pierwszego wykorzystania stworzonego profilu. Obowiązek ten aktualizuje na nowo się przy każdorazowej zmianie celu przetwarzania danych, nawet uprzednio posiadanych przez administratora. W przypadku danych uzyskanych pośrednio informacje te należy podać z zachowaniem terminów określonych w art. 14 ust. 3 RODO. Informacje muszą być widoczne i zrozumiałe, tj. przedstawione prostym językiem. Tak określony mechanizm ochrony praw osoby, której dane są przetwarzane w ASD, należy kwalifikować przede wszystkim jako konsekwencję zasady przejrzystości[367], minimalizacji danych[368] oraz ograniczenia celu przetwarzania, który powinien być zawsze wyraźny, prawnie uzasadniony i wiadomy osobie, której dane dotyczą[369].

Celem zasady przejrzystości jest umożliwienie podmiotom danych realnego dostępu do wiedzy oraz kontroli nad zakresem i kontekstem przetwarzania danych osobowych. Realizacja tego założenia już na początkowym etapie stosowania przepisów ochrony danych osobowych napotyka jednak na istotne utrudnienia. Przede wszystkim brakuje jasności co do tego, jak ustalić adresata obowiązków wynikających z RODO. Korzystanie z automatycznych systemów decyzyjnych poprzedzone jest bowiem złożonym i wieloetapowym procesem obejmującym projektowanie, uczenie algorytmów i wdrażanie rozwiązań w organizacjach[370]. Uczestnicy procesu mogą zatem mieć obowiązki zarówno na etapie fazy projektowania danego systemu, jak i na etapie wykorzystywania ASD. Nie jest przy tym oczywiste, czy status administratora

[365] Art. 35 RODO.
[366] *Google LLC vs. Commission nationale de l'informatique et des libertés (CNIL)*, 2019, https://privacylibrary.ccgnlud.org/case/google-llc-vs-commission-nationale-de-linformatique-et-des-liberts-cnil (dostęp: 4.12.2021).
[367] Art. 5 ust. 1 lit. a) RODO.
[368] Art. 5 ust. 1 lit. c) RODO.
[369] *Ogólne rozporządzenie o ochronie danych osobowych*, red. M. Sakowska-Baryła, Warszawa 2018, preprint 22, nb. 1.
[370] A. Krasuski, *op. cit.*

104:41105

należy przypisać bezpośrednio operatorowi systemu, czy może dostawcy rozwiązania informatycznego.

Administrator zgodnie z definicją powinien bowiem móc samodzielnie lub wspólnie z innymi decydować o celach i sposobach przetwarzania danych, a zatem obowiązki adresowane są do podmiotu, który określa odpowiednio, dlaczego i jak prowadzi się pewne czynności. Taka definicja obowiązanego nie przystaje do rzeczywistego zakresu ludzkiego nadzoru i kontroli automatycznych procesów decyzyjnych. Zastosowanie ASD, szczególnie tych opartych na algorytmach samouczących i wykorzystujących głębokie uczenie, powoduje bowiem, że operacje na danych mogą mieć charakter nieprzewidywalny i niewytłumaczalny zarówno dla projektanta, operatora, jak i użytkownika systemu. Wobec występowania barier poznawczych wypełnienie obowiązków informacyjnych i rzetelna realizacja praw podmiotów danych w przypadku ASD będzie czasami utrudniona czy wręcz uniemożliwiona. Nawet przy realizacji pełnego obowiązku informacyjnego można mieć wątpliwości, czy przejrzystość zaradzi sytuacji różnicowania w wypadku procesów, nad którymi jednostki i operatorzy nie mają realnej kontroli[371].

Należy również wskazać, że środowisko, w którym ASD gromadzi dane osobowe, może być kształtowane przez podmioty trzecie, które mogą mieć wpływ na zarządzanie zasobami sieciowymi, jak też na dostępność danych i sposób ich prezentacji[372]. Zaawansowane systemy samodzielnie określają sposoby gromadzenia danych, co ostatecznie może także prowadzić do modyfikacji założonych celów i sposobów przetwarzania danych. Problem ten staje się jeszcze bardziej istotny przy uwzględnieniu usług świadczonych przez systemy przetwarzające w chmurach obliczeniowych[373]. Identyfikacja ludzkiego administratora na obecnym etapie jest co prawda obarczona istotnym ryzykiem błędu, jednak konieczna. Systemy SI nie mają bowiem podmiotowości prawnej, na skutek czego SI nie mogą być wprost adresatem przepisów.

Wątpliwości dotyczące wskazania adresata obowiązków nie są jedynym utrudnieniem stosowania przepisów RODO do systemów automatycznych wykorzystujących algorytmy, w tym algorytmy sztucznej inteligencji. Zgodnie z zasadą przejrzystości w wypadku podejmowania automatycznych decyzji z wykorzystaniem profilowania podmiot danych powinien otrzymać: 1. istotne informacje o założeniach przetwarzania (zasadach przetwarzania) oraz 2. informacje o znaczeniu i przewidywanych konsekwencjach

[371] S. Wachter, *op. cit.*, s. 9.
[372] A. Krasuski, *Status prawny*…
[373] A. Krasuski, *Chmura obliczeniowa. Prawne aspekty zastosowania*, Warszawa 2017, s. 292.

105:87733

przetwarzania[374]. Realizacja tych obowiązków w kontekście algorytmów uczenia maszynowego może być z zasady niezupełna, a czasami nawet niemożliwa do spełnienia.

Dokonując porównania treści motywu 71 RODO oraz przepisu art. 22 RODO, odnoszących się do zagadnienia profilowania, należy ponadto wskazać na następującą niespójność ustawodawcy unijnego. W motywie 71 RODO ustawodawca przyjął m.in., że osoba fizyczna powinna mieć „prawo do uzyskania wyjaśnienia co do decyzji wynikłej z takiej oceny", przy czym tego uprawnienia nie uwzględnił w art. 22 ust. 3 RODO, w którym określił uprawnienia osób fizycznych związane z profilowaniem, ani też w żadnym innym przepisie RODO[375]. Prawo do wyjaśnienia jest zatem z jednej strony niedoprecyzowane w treści samej regulacji, z drugiej zaś pojawiają się problemy z implementacją regulacji do obecnych rozwiązań informatycznych stosowanych w omawianych w niniejszej pracy systemach. Nawet twórcy ASD mogą bowiem nie mieć wystarczających kompetencji i narzędzi umożliwiających jednoznaczne wyjaśnienie sposobu działania systemu, algorytmów lub modeli SI, albowiem mogą się one stale rozwijać i zmieniać[376].

Zasada przejrzystości rozumiana w tak szerokim ujęciu może przy tym konkurować z innymi dobrami chronionymi, jak np. swoboda gospodarcza, tajemnica przedsiębiorstwa lub własność intelektualna. Dostrzegając to ryzyko, w motywie 63 RODO zapewniono pewien stopień ochrony administratorów, wskazując, że prawo dostępu „nie powinno negatywnie wpływać na prawa lub wolności innych osób, w tym tajemnice handlowe lub własność intelektualną, w szczególności na prawa autorskie chroniące oprogramowanie ASD". Wobec tego należy przyjąć, że przekazywane podmiotom danych informacje powinny być wystarczające do rzeczywistego umożliwienia skorzystania przez podmioty danych z przysługujących im praw[377], przy czym nie będą one obejmowały konieczności wyjaśnienia działania algorytmu[378].

Realizacja nawet tak okrojonego obowiązku będzie jednak wystarczająco utrudniona z uwagi na specyfikę uczenia maszynowego. W procesie analizy danych mogą bowiem ulegać zmianie zarówno pierwotnie ustalone sposoby przetwarzania, jak i zasady podejmowania decyzji. W

[374] Art. 22 RODO.
[375] A. Burt, *Is There a "Right to Explanation" for Machine Learning in the GDPR?*, IAPP 2017, https://iapp.org/news/a/is-there-a-right-to-explanation-for-machine-learning-in-the-gdpr/ (dostęp: 4.12.2021).
[376] D. Lubasz, M. Namysłowska, *op. cit*.
[377] *Ibidem*.
[378] Grupa Robocza Art. 29 w wytycznych dotyczących zautomatyzowanego wydawania decyzji z 2017 r. wskazała, że prawo do informacji jest raczej bardziej ogólną formą nadzoru, a nie prawem do uzasadnienia konkretnej decyzji.

106:64311

konsekwencji również zakres oddziaływania automatycznych decyzji przy wykorzystaniu profilowania na podmioty danych może się zmieniać.

Człowiek nie jest przy tym przygotowany i zdolny do pełnego korzystania z praw przysługujących mu na podstawie RODO. Wynika to nie tylko z nieprzejrzystości i niejawności procesów decyzyjnych, ale też z braku dostatecznej świadomości skali wykorzystania danych osobowych w związku z pożytkowaniem urządzeń końcowych, stosowanych permanentnie jako urządzenia wchodzące w skład IoT[379]. W związku z tym aplikowanie przepisów RODO do automatycznych systemów decyzyjnych doznaje istotnych ograniczeń ze względu na przedmiot regulacji i specyfikę istniejących rozwiązań technologicznych. Opisane wymogi nie znajdą przy tym żadnego zastosowania, jeżeli ASD będzie wykorzystywany do przetwarzania danych osobowych, ale nie będzie służył ani do profilowania, ani do całkowicie zautomatyzowanego podejmowania decyzji w rozumieniu RODO. Ewentualne obowiązki będą wynikać jedynie z ogólnych zasad RODO, tj. rzetelności i przejrzystości, jednak i one będą w swoim zakresie niedookreślone[380].

Oceniając zakres zastosowania wskazanej regulacji do problemu różnicowania przez automatyczne systemy decyzyjne, dochodzę do wniosku o niezapewnieniu odpowiedniej ochrony jednostki na gruncie przywołanych przepisów. Negatywna ocena ochrony przed różnicowaniem algorytmicznym na gruncie RODO wynika z fragmentaryczności regulacji, braku realnych kompetencji administratora danych do realizacji obowiązków informacyjnych, braku narzędzi do nadzoru i kontroli nad prawidłowością wykonywania tychże obowiązków, a także koncentracji regulacji na procesie przetwarzania, a nie jego skutkach. Reasumując, należy przyjąć, że RODO jest nieprzystosowane, żeby swoją regulacją objąć problem różnicowania w toku stosowania ASD, w tym przy wykorzystaniu algorytmów sztucznej inteligencji. Nie można zatem tego aktu normatywnego uznawać za wystarczające źródło prawa do ochrony osób fizycznych w kontekście zagrożeń automatycznej dyskryminacji. Teza ta znajduje potwierdzenie w literaturze przedmiotu[381].

[379] A. Krasuski, *Status prawny*…

[380] K. Syska, *Ocena odpowiedniości przepisów RODO do zapewnienia przejrzystości działania systemów AI – wybrane zagadnienia*, [w:] *Ocena i przegląd RODO po dwóch latach obowiązywania*. *Aktualne problemy prawnej ochrony danych osobowych 2020*, red. G. Sibiga, dodatek specjalny do „Monitora Prawniczego" 23, 2020, s. 1577.

[381] R. van den Hoven van Genderen, *Do We Need New Legal Personhood in the Age of Robots and AI?*, [w:] *Robotics, AI and the Future of Law*, red. M. Corrales, M. Fenwick, N. Forgó, Singapore 2018, s. 15–50.

107:76503

Przedstawione niedoskonałości dotychczasowych regulacji dotyczących ochrony danych osobowych przy jednoczesnym rosnącym zapotrzebowaniu i uzależnieniu rozwoju ASD, w tym SI, od dostępności i wiarygodności danych są motorem dalszych prac legislacyjnych na gruncie prawa unijnego. Parlament Europejski 6.04.2022 r. w ramach uzupełnienia RODO przyjął kolejny akt prawny, wspierający budowanie ram godnej zaufania sztucznej inteligencji, tj. rozporządzenie o zarządzaniu danymi (ang. *Data Governance Act*)[382]. Przepisy tej regulacji mają na celu poprawę jakości danych poprzez umożliwienie ponownego wykorzystania danych zbieranych przez sektor publiczny. W efekcie ma to sprawić, że obrót danymi stanie się bezpieczniejszy i łatwiejszy, a także zapewnić zgodność z przepisami o ochronie danych. Umożliwienie szerszego ponownego wykorzystania danych będących w posiadaniu organów sektora publicznego i utworzenie jednolitego zharmonizowanego rynku obrotu danymi pozwoli też zapewnić wyższy poziom zaufania do danych, a przez to i technologii opierającej się na ich analizie.

Warunki ponownego wykorzystywania danych powinny być przy tym niedyskryminujące, proporcjonalne i obiektywnie uzasadnione, a także nie mogą ograniczać konkurencji[383]. Zakres rozporządzenia ma objąć m.in. tajemnice handlowe, dane osobowe i dane chronione prawami własności intelektualnej[384], przy czym organy sektora publicznego umożliwiające tego rodzaju ponowne wykorzystanie będą musiały być odpowiednio wyposażone pod względem technicznym, aby zapewnić pełne zachowanie prywatności i poufności. W tym celu należy rozważyć wykorzystanie bezpiecznych środowisk przetwarzania danych i technik anonimizacji, takich jak prywatność różnicowa i tworzenie danych syntetycznych.

Wprowadzono już system licencjonowania „pośredników danych"[385], tj. organizacji, które ustanawiają porozumienia handlowe między posiadaczami danych a ich użytkownikami. Rozporządzenie zawiera ponadto zarys ram ograniczenia przekazywania danych nieosobowych, które mają być chronione przed bezprawnym lub nieuprawnionym dostępem na mocy prawa unijnego lub krajowego. Jeśli takie dane znajdują się w posiadaniu organów sektora publicznego, powinny być przekazywane wyłącznie do państw trzecich, w których zapewnione są odpowiednie zabezpieczenia w zakresie wykorzystywania danych. W założeniu poprawie ma dzięki temu ulec efektywność kontroli, jaką podmioty danych uzyskają nad

[382] Wniosek: rozporządzenie Parlamentu Europejskiego i Rady w sprawie europejskiego zarządzania danymi (akt w sprawie zarządzania danymi), COM(2020) 767 final.
[383] Pkt 11 preambuły COM(2020) 767 final.
[384] Art. 3 COM(2020) 767 final.
[385] Art. 9–10 COM(2020) 767 final.

108:37045

sposobem, w jaki udostępniają swoje dane (planowane jest wprowadzenie przestrzeni danych osobowych oraz portfeli danych). Wskazane instrumenty mają również stworzyć warunki sprzyjające „altruizmowi w zakresie danych”[386], czyli zachęcaniu osób fizycznych do dobrowolnego przekazywania swoich danych osobowych w interesie ogólnym.

RODO jako regulacja odpowiadająca na podobny typ niewidocznego i inherentnego ryzyka naruszenia praw i wolności na skutek przetwarzania danych powinno przynajmniej identyfikować zagrożenia związane z niedozwolonym różnicowaniem przez automatyczne systemy decyzyjne. Należy w tym zakresie pozytywnie ocenić kierunki regulacyjne, ale jednocześnie wprost zaznaczyć ograniczony zakres ochrony przez RODO przed automatyczną dyskryminacją. Wyzwania te uzasadniają potrzebę analizy i wykładni przepisów o ochronie danych z perspektywy nowych zagrożeń związanych z ASD w kontekście nieprzejrzystości i masowości procesów różnicowania. Szkodliwe różnicowanie algorytmiczne może się bowiem wymykać instrumentom penalizującym naruszenia ochrony danych, gdyż niekoniecznie procesy prowadzące do dyskryminacji będą naruszały zasady dotyczące danych osobowych[387]. Obecnie wdrażane wysokie standardy ochrony danych osobowych nie wykluczają zatem konieczności prawidłowego zdefiniowania, a następnie zmitygowania ryzyk dotyczących rozwoju technologii opartych na rozwiązaniach SI[388].

4. Prawo do prywatności

Jednym z kluczowych elementów, których prywatność musi zostać zachowana, są dane i informacje chronione przed publicznym dostępem. Przetwarzanie dużych zbiorów danych w ASD i oparte na nich wnioskowanie może zaś być narzędziem do masowej inwigilacji i predykcji, jak również stać się podstawą argumentacji w kwestii dyskryminacji statystycznej. Upowszechnienie analizy dużych zbiorów danych, stanowiące esencję funkcjonowania algorytmów, prowadzi zatem do inherentnego ryzyka naruszenia praw i wolności jednostki, które – jak wskazałam w poprzednim podrozdziale – może się wymykać obowiązującym regulacjom czy klasycznym instrumentom władzy i nadzoru. Niejawne procesy gromadzenia, łączenia i innych form przetwarzania danych, wzmacniane nieprzejrzystością epistemiczną

[386] Rozdz. IV COM(2020) 767 final.
[387] Zob. wyrok TSUE (wielka izba) z 16.12.2008 r. w sprawie *Heinz Huber v. Bundesrepublik Deutschland*, skarga nr C-524/06.
[388] D. Lubasz, M. Namysłowska, *op. cit.*

109:71136

ASD, mogą także – jak w przypadku SyRI – masowo wspierać i umacniać marginalizację grup historycznie wykluczonych.

Samuel D. Warren i Louis D. Brandeis[389] wskazywali, że rozwój technologii i konieczności ochrony sfer życia narażonych w tym kontekście na ich oddziaływanie uzasadnia wyodrębnienie i stosowanie prawa do prywatności. Pod tym względem można zauważyć analogię między systemami automatycznymi a innymi procesami inwigilacji spotykanymi na gruncie wykonywania władzy publicznej. Zastosowanie gwarancji prawa do prywatności chociażby uzupełniająco, tj. w zakresie różnicowania przez ASD na podstawie danych niebędących cechami chronionymi prawem antydyskryminacyjnym czy poza zakresem regulacji RODO, umożliwiłoby ograniczenie negatywnego wpływu różnicowania algorytmicznego na prawa i obowiązki jednostki.

Wobec niekoherentności pojęcia prawa do prywatności na potrzeby niniejszej analizy przyjęłam za częścią doktryny[390], że prywatność należy rozumieć jako prawo kontroli wykorzystania informacji o samym sobie, czyli prawo decydowania o zakresie i zasięgu udostępniania oraz komunikowania innym osobom informacji o swoim życiu[391], oraz prawo do tego, aby jednostka w pewnym obszarze pozostawała wolna od ingerencji[392]. Zgodnie z poglądem Luciana Floridiego sama tożsamość, a także informacje ją tworzące mają wartość niezależną od roli, jaką odgrywają w podejmowaniu decyzji, które przynoszą szkodliwe lub korzystne skutki.

Integralność tożsamości podmiotu zostaje naruszona, gdy dane lub informacje są dodawane do tożsamości podmiotu bez jego zgody[393]. Sfera życia prywatnego jest przy tym dobrem osobistym zawierającym wszystko to, co ze względu na uzasadnione odosobnienie się jednostki od ogółu społeczeństwa służy jej m.in. do osiągnięcia i zachowania oczekiwanej pozycji społecznej[394]. Prawo do prywatności może zatem zarówno zabezpieczać autonomię informacyjną jednostki, jak i chronić ją przed dowolnym zestawianiem oraz analizą danych i

[389] S.D. Warren, L.D. Brandeis, *The Right to Privacy*, „Harvard Law Review" 4, 1890, nr 5, s. 193–220.
[390] I. Gil *et al.*, *Konwencja o ochronie praw człowieka i podstawowych wolności z dnia 4.11.1950*, red. Ł. Błaszczak, t. 2, Legalis/el. 2021, art. 8.
[391] Zob. wyrok TK z 24.06.1997 r., sygn. K 21/96, OTK-A 1997, Nr 2, poz. 23; wyrok TK z 19.05.1998 r., sygn. U 5/97, OTK 1998, Nr 4, poz. 46; wyrok TK z 27.01.1999 r., sygn. K 1/98, ZU 1999, Nr 1, poz. 3.
[392] L. Kański, *Prawo do prywatności*, [w:] *Prawa człowieka. Model prawny*, red. R. Wieruszewski, Warszawa 1991, s. 329.
[393] Prywatność grupowa opiera się na koncepcji tożsamości informacyjnej Luciana Floridiego; zob. *idem*, *Open Data, Data Protection, and Group Privacy*, „Philosophy & Technology" 27, 2014, nr 1, s. 1–3.
[394] A. Kopff, *Koncepcja praw do intymności i do prywatności życia osobistego*, „Studia Cywilistyczne" 20, 1972, s. 32–33.

110:90230

informacji posiadanych przez państwo i podmioty komercyjne w swoich systemach. Należy się zgodzić z Dominikiem Lubaszem, że:

> w świecie online leżąca u podstaw ochrony naszych danych koncepcja autonomii informacyjnej jest istotnie ograniczona poprzez sposób nie tyle, lub nie tylko, dzielenia się informacjami, ale sposób zbierania i opracowania informacji przez uczestników cyfrowego obrotu – wydawców, reklamodawców, pośredników i brokerów czy wreszcie portale społecznościowe, agencje reklamowe i dostawców usług analitycznych[395].

W zakresie tego prawa znajdzie się więc zarówno tworzenie na podstawie analizy danych profili indywidualnych, jak i zestawianie takich wyników analizy z profilami grupowymi w celu odkrywania informacji nieudostępnionych przez jednostkę czy też poszukiwanie szerszych korelacji i uogólnień. Dane nie tylko są bowiem zbierane od ich dysponentów, ale są także odkrywane na skutek wnioskowania o regułach przez algorytmy, w tym algorytmy predykcyjne[396]. Na podstawie tak zebranych danych i ustalonych zależności pomiędzy nimi operatorzy automatycznych procesów decyzyjnych mogą istotnie wpływać na prawa i obowiązki podmiotu danych, w tym przyczyniać się do tworzenia niedokładnych lub nieprawidłowych kategorii i kwalifikacji.

Prawo do prywatności współistnieje przy tym z zakazem dyskryminacji i regulacjami dotyczącymi danych osobowych, a jednym z jego celów jest ochrona przed stygmatyzacją czy stereotypami. Prawo to uzupełnia również omówione do tej pory przepisy, zakazując m.in. różnicowania, które może naruszać prywatność jednostki. Będzie ono mogło mieć też istotne znaczenie dla ochrony praw jednostki w sytuacjach, w których nie zostaną spełnione przesłanki RODO, lecz także gdy zaistnieją formalne przeszkody w wykazaniu naruszeń na podstawie wskazanej regulacji. Takie sytuacje mogą mieć miejsce w wypadku różnicowania opartego na analizie danych nieosobowych, podlegających co prawda pod regulację RODO, jednak wymykających się dodatkowym obowiązkom informacyjnym ustanowionym przez art. 22 RODO, jak również na danych i informacjach wyinterpretowanych z danych niepodlegających tej regulacji.

W krajowym systemie prawnym prawo do prywatności znajduje uzasadnienie w porządku konstytucyjnym, ale też prywatnoprawnym. Artykuł 47 Konstytucji RP formułuje wprost prawo do „ochrony prawnej życia prywatnego, rodzinnego, czci i dobrego imienia oraz do

[395] D. Lubasz, Big Brother…
[396] W skład zbieranych danych wchodzą: dane udostępniane przez podmioty danych, dane zaobserwowane, dane behawioralne czy dane wyinterpretowane przy wykorzystaniu algorytmów; por. *ibidem*.

111:85556

decydowania o swoim życiu prywatnym" (zbiorczo określane właśnie jako prawo do prywatności), przy czym przepis ten jest uzupełniany licznymi dodatkowymi regulacjami, które są emanacją prawa do prywatności[397]. Treść prawa do prywatności można np. wywieść z art. 17 MPPOiP[398] oraz art. 12 PDPCz. Z kolei w orzecznictwie Trybunału Konstytucyjnego[399] prezentowane jest stanowisko, że nie jest możliwe precyzyjne wskazanie i enumeratywne wyliczenie elementów składowych prawa do prywatności. Cywilistycznie prawo do prywatności stanowi też przedmiot ochrony jako samoistne dobro osobiste człowieka[400].

Prawo do prywatności znajduje wyraz również na gruncie międzynarodowym. Europejska konwencja o ochronie praw człowieka i podstawowych wolności[401] nie wprowadza wprawdzie definicji legalnej pojęcia prywatności, jednak z uwagi na ścisły związek tego prawa z tożsamością osobistą dostrzega podobieństwa do ochrony danych osobowych. Prawo do prywatności zostało ponadto zagwarantowane w Karcie Praw Podstawowych[402]. Ingerencja w prawo do prywatności jest przy tym dopuszczalna tylko wtedy, gdy jest zgodna z prawem, dąży do legitymowanego celu i można ją uznać za konieczną w społeczeństwie demokratycznym.

Analizując aspekt prywatności związany z dostępem do informacji o jednostce, niezbędna jest analiza tego, w jakim stopniu informacja musi być powiązana z jednostką, aby ujawnienie tej informacji miało wpływ na utratę przez jednostkę prywatności[403].

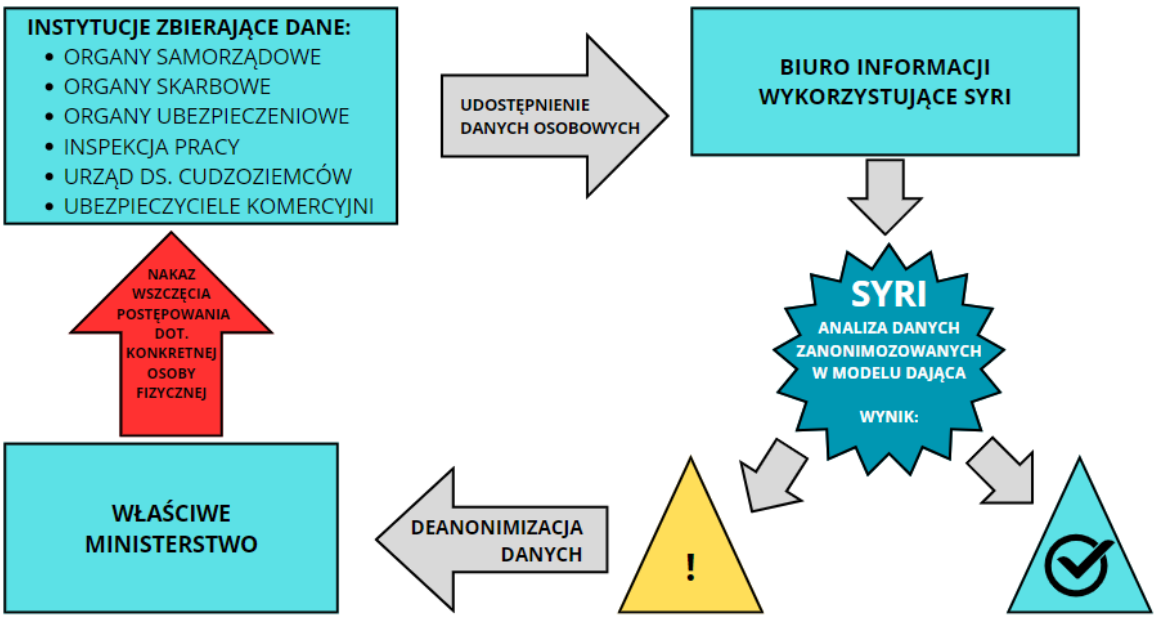
Niestety systemy automatyczne osiągnęły poziom zaawansowania i nieprzejrzystości w analizie danych, który użytkownikowi trudno sobie wyobrazić. Jednostka nie tylko nie wie, na jakich danych wyszkolony został ASD, ale również jakie kategorie i informacje system uczynił decydującymi dla konkretnego procesu decyzyjnego. Przykładem zastosowania prawa do prywatności w celu ograniczenia ASD była sprawa wspominanego już wielokrotnie SyRI. Ten niderlandzki system powstał, by zapobiegać i zwalczać nadużycia w dziedzinie ubezpieczeń społecznych i systemów uzależnionych od dochodów, podatków i ubezpieczeń społecznych oraz prawa pracy. W rzeczywistości było to narzędzie informatyczne przetwarzające w różnych formach dane zgromadzone w państwowych rejestrach zatrudnienia, systemach

[397] Art. 48–51 Konstytucji RP.
[398] Międzynarodowy Pakt Praw Obywatelskich i Politycznych otwarty do podpisu w Nowym Jorku 19.12.1966 r., Dz.U. z 1977 r. Nr 38, poz. 167.
[399] Wyrok TK z 27.01.1999 r., sygn. K 1/98, ZU 1999, Nr 1, poz. 3; wyrok TK z 13.07.2004 r., sygn. K 20/03, OTK-A 2004, Nr 7, poz. 63.
[400] Art. 23 kodeksu cywilnego, Dz.U. z 1964 r. Nr 16, poz. 93 (dalej: KC).
[401] Art. 8 EKPCz.
[402] Art. 7 KPP.
[403] R. Gavison, *Privacy and the Limits of Law*, „The Yale Law Journal" 89, 1980, nr 3, s. 421–471.

112:20498

świadczeń socjalnych, raportach z instytucji kredytowych, historii edukacji i zameldowania. Na podstawie tych danych system tworzył profile i kategoryzował obywateli pod kątem ryzyka popełnienia przestępstwa. Był jednak wykorzystywany do badania dzielnic emigranckich, które zostały uznane za obszary problemowe, w związku z czym wzbudził zainteresowanie, a następnie wywołał powszechną krytykę związaną z zarzutem dyskryminacji ubogich, wykluczonych i mniejszości narodowych.

Z perspektywy administracji publicznej system dokonywał dokładnej predykcji, dzięki czemu miał zwiększać szanse na wykrycie nieprawidłowości. SyRI skupiał się przy tym na dzielnicach emigranckich, ponieważ zadziałał mechanizm sprzężenia zwrotnego algorytmu. Wizerunek dzielnicy jako obszaru problemowego przyczyniał się jednak do powstawania stereotypów i wzmacniania negatywnego obrazu mieszkańców takich dzielnic, nawet jeśli przestępczość w tych rejonach była statystycznie porównywalna do innych rejonów. Koncentracja analizy danych pochodzących z danego obszaru doprowadziła zatem do nieproporcjonalnych i niereprezentatywnych wyników.



Rys. 3. Schemat blokowy pokazujący, jak działał SyRI

Źródło: *Q&A: How the EU's Flawed Artificial Intelligence Regulation Endangers the Social Safety Net*, https://www.hrw.org/news/2021/11/10/how-eus-flawed-artificial-intelligence-regulation-endangers-social-safety-net (dostęp: 4.12.2021).

113:92664

Przechodząc do szczegółowego omówienia samego postępowania w sprawie SyRI, należy wskazać, że zostało ono wszczęte przez koalicję organizacji społeczeństwa obywatelskiego. Sprawa funkcjonowania systemu SyRI zawisła w marcu 2018 r. przed Sądem Okręgowym w Hadze przeciwko rządowi niderlandzkiemu. W toku procesu Królestwo Niderlandów wskazywało, że system SyRI ma oparcie w przepisach prawa, jako że legalizowała go ustawa zezwalająca na stosowanie algorytmów głębokiego uczenia i eksplorację dużych zbiorów danych. Sam kod algorytmu, jak również bazy danych czy szczegółowe informacje o sposobie funkcjonowania systemu nie zostały jednak ujawnione. Jako uzasadnienie podano ochronę przed ewentualną próbą obejścia algorytmu przez oszustów ubezpieczeniowych. Sąd uznał w związku z tym, że nie można było jednoznacznie zweryfikować, czy SyRI używa algorytmu dyskryminacyjnego. Jednocześnie brak przejrzystości algorytmu w zakresie sposobu określenia ryzyka naruszenia prawa przez obywatela uniemożliwiał ludziom kwestionowanie rozstrzygnięć SyRI, co z kolei naruszało prawo do prywatności[404].

Chociaż sąd nie stwierdził, że dyskryminacja rzeczywiście miała miejsce, ponieważ nie posiadał odpowiednich narzędzi regulacyjnych, doszedł do wniosku, że ryzyko dyskryminacji w połączeniu z brakiem przejrzystości świadczy o naruszeniu art. 8 EKPCz. Tym sposobem uznał różnicowanie wymykające się klasycznej definicji dyskryminacji za naruszające prywatność. W ocenie sądu państwo ponosi bowiem szczególną odpowiedzialność przy stosowaniu nowych technologii, która koniecznie wymaga określenia równowagi między korzyściami, jakie ich stosowanie przynosi w zakresie zapobiegania i zwalczania nadużyć finansowych z jednej strony a potencjalną ingerencją w wykonywanie prawa do poszanowania życia prywatnego poprzez takie rozwiązania z drugiej. Sąd nie znalazł jednak podstaw do zastosowania wprost art. 14 EKPCz, który zakazuje dyskryminacji. Nieprzeprowadzenia analizy dotyczącej możliwego zastosowania zakazu dyskryminacji wobec różnicowania algorytmicznego jedynie z powodu nieprzejrzystości algorytmu i nieudzielenia informacji o jego funkcjonowaniu nie sposób przy tym ocenić pozytywnie.

Ostatecznie 5.02.2020 r. sąd w Hadze nakazał natychmiastowe wstrzymanie stosowania SyRI. Sąd stwierdził m.in., że organ administracji nie był w stanie właściwie uzasadnić działań i motywów automatycznych swoich decyzji. Operator systemu automatycznego nie rozumiał bowiem, dlaczego system stwierdzał istnienie związku między danymi a predykcją. Z tej przyczyny ustalono, że algorytm SyRI naruszał prawo do poszanowania życia prywatnego i

[404] A. Toh, *Dutch Ruling a Victory for Rights of the Poor*, Human Rights Watch 2020, https://www.hrw.org/news/2020/02/06/dutch-ruling-victory-rights-poor (dostęp: 4.12.2021).

114:11381

rodzinnego. Co więcej, sąd uznał, że ustawodawstwo legalizujące SyRI nie zawierało wystarczających zabezpieczeń, aby stwierdzić, iż było w takim kształcie konieczne w społeczeństwie demokratycznym, czego wymaga art. 8 ust. 2 EKPCz[405]. Królestwo Niderlandów nie odwołało się od orzeczenia i zaprzestało korzystania z SyRI w jej ówczesnym kształcie.

Należy jednak zaznaczyć, że omawiana sprawa nie była pierwszym sporem na gruncie EKPCz dotyczącym naruszenia prywatności przez automatyczne systemy decyzyjne. Podobnie orzekł Europejski Trybunał Praw Człowieka w sprawie węgierskiego systemu nadzoru elektronicznego[406].

Analogię można dostrzec również w polskiej sprawie związanej z kwestionowaniem uprawnienia do stosowania narzędzia do profilowania bezrobotnych. Rzecznik Praw Obywatelskich Adam Bodnar w swoim wniosku do Trybunału Konstytucyjnego wskazał bowiem, że procedura profilowania jako czynność polegająca na przetworzeniu jakiejś ilości informacji o osobie prowadzi do wniosku, iż ASD ogranicza prawo do prywatności i autonomii informacyjnej osób bezrobotnych[407].

Nie można też tracić z pola widzenia, że profilowanie – tak jak każde działanie organów administracji wobec jednostki – powinno być rzetelne i sprawne, odbywać się na podstawie i w granicach prawa oraz z poszanowaniem zasady równości i praw podmiotowych, w tym m.in. prawa do prywatności i autonomii informacyjnej[408].

W przypadku opisywanego systemu do profilowania podmiot danych nie był uprawniony do zakwestionowania wyniku analizy. Polski Trybunał Konstytucyjny nie dostrzegł jednak podstaw, aby zaskarżone przepisy zakwestionować z perspektywy naruszenia prawa do prywatności[409]. Co więcej, w orzeczeniu[410] stwierdził, że gromadzenie danych, wpisywanie ich do kwestionariusza, a także ich przetwarzanie oraz ustalanie na ich podstawie profilu nie przesądza w sposób władczy o sytuacji prawnej danego podmiotu ani w relacji z innymi równorzędnymi podmiotami, ani w relacji z władzą publiczną. Dlatego też Trybunał nie

[405] Trybunał zastosował zatem ogólne zasady ochrony danych z art. 5 RODO w celu weryfikacji spełnienia wymogów stawianych przez art. 8 EKPCz, a dokładniej kryterium, zgodnie z którym ingerencja państwa w prywatność powinna być „konieczna w demokratycznym społeczeństwie".
[406] Wyrok ETPCz z 12.01.2016 r. w sprawie *Szabó and Vissy v. Hungary*, skarga nr 37138/14.
[407] Wniosek Rzecznika Praw Obywatelskich, sygn. VII.501.277.2014.KM.
[408] Por. preambuła, art. 2, art. 7, art. 32 ust. 1, art. 47 i art. 51 Konstytucji oraz wyrok TK z 6.06.2018 r., sygn. K 53/16, Dz.U. poz. 1149.
[409] Wyrok TK z 6.06.2018 r., sygn. K 53/16.
[410] *Ibidem*.

115:51745

zakwalifikował postępowania, które łączy się z ustaleniem profilu, za konstytuujące „sprawę" w świetle art. 45 ust. 1 Konstytucji. W uzasadnieniu orzeczenia podano, że profilowanie jest prawną formą działania administracji *sui generis*, polegającą na diagnozowaniu sytuacji bezrobotnego i wyborze pakietu najbardziej odpowiednich dla danej osoby form pomocy, a zatem jako czynność materialno-techniczna etap ten nie musi być objęty gwarancjami wynikającymi z art. 45 ust. 1 i art. 78 Konstytucji[411].

Trybunał uznał również, że nawet druga faza tej czynności (tj. przeliczenie informacji o bezrobotnym na punkty i ustalenie na tej podstawie odpowiedniego profilu) to jedynie kolejne przybliżenie do udzielenia konkretnej formy pomocy, wykluczające instrumenty nieuwzględnione w danym profilu, a równocześnie nieotwierające drogi do bezwarunkowego żądania uzyskania każdego ze środków w nim przewidzianych. Profilowanie w ocenie TK nie było więc równoznaczne z ostatecznym władczym rozstrzyganiem o prawach lub obowiązkach jednostek, a tylko stwarzało warunki do wydania takich rozstrzygnięć jako jedna z okoliczności branych pod uwagę przez urząd pracy. Trybunał pominął jednak zarzuty związane z brakiem realnego nadzoru i weryfikacji ludzkiej tak wykreowanych automatycznych prognoz.

Ostatecznie w uzasadnieniu orzeczenia wskazano, że ustawodawca w ramach przysługującej mu swobody regulacyjnej mógł przyznać pierwszeństwo efektywności i sprawności działania administracji kosztem prawa do kwestionowania wszystkich czynności poprzedzających wydanie konkretnych rozstrzygnięć dotyczących bezrobotnego, aczkolwiek Trybunał stwierdził, że art. 34a ust. 3c ustawy o promocji zatrudnienia jest niezgodny z art. 92 ust. 1 Konstytucji z powodu braku wytycznych merytorycznych co do treści przewidzianego w nim rozporządzenia. Kształt profilowania (tj. kryteria i procedury) nie mógł być zatem w ocenie TK ustalany samoistnie w rozporządzeniu oraz narzędziach stosowanych przez urzędy prac.

W efekcie orzeczenia system do profilowania przestał działać, jednak Trybunał pozostawił wadliwe rozwiązanie na dalsze 12 miesięcy w celu przygotowania i wprowadzenia nowych ram prawnych.

Omawiany wyrok ukazuje niestety ograniczoną refleksję nad społecznymi skutkami automatycznych narzędzi do profilowania, a jednocześnie w swoich ustaleniach nie przystaje do aktualnej roli, jaką ASD odgrywa w kształtowaniu praw i obowiązków obywateli. Trybunał

[411] M. Wyrzykowski, M. Ziółkowski, *Konstytucyjny status proceduralny jednostki jako adresata działań organów administracji*, red. R. Hauser, A. Wróbel, Z. Niewiadomski, „System Prawa Administracyjnego", t. 2, Warszawa 2012, s. 283 oraz wyrok pełnego składu TK z 22.06.2015 r., sygn. SK 29/13, OTK-A 2015, Nr 6, poz. 83, cz. III, pkt 7.8 i 7.5 – uzasadnienia.

116:86211

nie odniósł się bowiem do ryzyk opisywanych przez organizacje pozarządowe i ekspertów związanych z marginalizacją i wykluczeniem osób w najmniej korzystnej sytuacji na rynku pracy. Bagatelizowanie skali zagrożenia i masowości problemu algorytmicznego różnicowania nie jest pozytywnym kierunkiem orzeczniczym i rodzi obawy o krajowe podejście do problemu automatycznej dyskryminacji.

Wszystkie przywołane orzeczenia wskazują jednak, że prawo do prywatności może pomocniczo służyć ograniczeniu skutków nieprawidłowego różnicowania przez ASD. Regulacja ta wydaje się pewnym pomostem lub uzupełnieniem wykorzystywanym przez sądy do uchwycenia niesprawiedliwości. Sądy potwierdzają przy tym, że odmienności dyskryminacji automatycznej względem klasycznego zakresu ochrony antydyskryminacyjnej czynią istniejące narzędzia prawne nieadekwatnymi. Próba adaptacji prawa do specyfiki automatycznych systemów decyzyjnych w drodze sporów sądowych jest dodatkowo podatna na różne szkodliwe bodźce – sędziowie zazwyczaj podejmują decyzje pod presją czasu, mając ograniczone informacje na temat rozwiązań technologicznych i szerszych konsekwencji ich stosowania, podczas gdy ustawodawca może zlecić przeprowadzenie badań i obiektywnych analiz przy opracowywaniu przepisów[412]. Pozostawienie ważnych decyzji społecznych wyłącznie sądownictwu z perspektywy alarmującego postępu technologicznego grozi powstaniem deficytu demokracji[413].

Opierając się na europejskim prawie antydyskryminacyjnym, można podsumować, że tego typu regulacje ze względu na zakres terytorialny i podmiotowy nie będą kompleksowo odpowiadały na problem różnicowania przez automatyczne systemy decyzyjne. Można przy tym zaobserwować próbę reakcji na odmienność różnicowania algorytmicznego od dotychczasowego sposobu podejmowania decyzji poprzez wykorzystanie reguł dotyczących ochrony prywatności czy ochrony danych osobowych. Regulacje te koncentrują się jednak na jednostkach, podczas gdy uzasadnieniem przeciwdziałania dyskryminacji jest szersza perspektywa społeczna. Ten drugi proces może bowiem nie tylko naruszyć prawa konkretnego obywatela, ale przede wszystkim doprowadzić do rozkładu społecznego. Co więcej, wraz ze wzrostem zastosowania algorytmów do procesów decyzyjnych uprzedzenia reprodukowane w systemach staną się bardziej wyrafinowane i trudne do zidentyfikowania, kontrolowania lub zakwestionowania. Fragmentaryczność unormowań w zakresie prawa zakazującego dyskryminacji, w tym brak wspólnego standardu w kwestii charakteru katalogu kryteriów

[412] J. Turner, *Robot Rules: Regulating Artificial Intelligence*, Cham 2019, s. 223.
[413] *Ibidem*, s. 224.

117:60924

niedozwolonego różnicowania oraz postępująca ewolucja definicji zasady równości i wielość koncepcji jej towarzyszących[414], będą mogły uczynić iluzorycznymi prawa jednostki w zakresie niepodlegania krzywdzącym skutkom automatycznej dyskryminacji. Te trudności są jedynie potęgowane deprecjonującym podejściem sądów i organów administracyjnych, które wyłącznie z powodu nieprzejrzystości systemów bagatelizują skutki masowości automatycznych systemów decyzyjnych.

Z zaprezentowanych w tym rozdziale rozważań wynika, że ani europejskie prawo antydyskryminacyjne w swoim aktualnym kształcie, ani RODO, ani też prawo do prywatności w sposób adekwatny i zupełny nie identyfikują i nie przeciwdziałają ryzykom marginalizacji grup historycznie wykluczonych na skutek stosowania algorytmów. Wobec dotychczasowego postępującego rozwoju ochrony przed dyskryminacją na tle wydarzeń społeczno-politycznych ukazanych w rozdziale I niniejszej pracy należy zatem postawić pytanie o możliwość dalszego rozwoju legislacji w celu skutecznego uregulowania nowego fragmentu prawa w zgodzie z postulatami równościowymi.

[414] M. Domańska, *op. cit.*, rozdz. 1.4.1.

118:54410

IV. Sztuczna inteligencja jako samodzielny przedmiot regulacji prawa Unii Europejskiej

Technologia staje się kolejnym frontem postulatów równościowych. Ryzyko naruszeń praw i wolności człowieka jest inherentne, przy czym mechanizmy je wzmacniające stają się niepojmowalne, niewidoczne i nieprzejrzyste dla użytkowników i operatorów systemów. W dodatku już na obecnym etapie rozwoju automatycznych systemów decyzyjnych uwidoczniają się negatywne skutki społeczne, które w sposób szczególny mogą dotykać osób należących do grup marginalizowanych. Jak bowiem wykazałam wcześniej, algorytmy są nieuchronnie obciążone wartościowaniem[415], a różnicowanie w toku automatycznego procesu decyzyjnego może mieć formę dyskryminacji, jak również – nie wypełniając prawnych przesłanek dyskryminacji – skutkować wykluczeniem i pogłębiać istniejące nierówności społeczne. Zagrożenia wynikające ze stosowania ASD istnieją także w obszarze publicznoprawnym, gdzie gwarancje równościowe nie powinny konkurować z postulatami wolnorynkowymi (por. SyRI czy polski ASD do profilowania bezrobotnych).

Z powodu niejawności i nieprzejrzystości procesów decyzyjnych ASD głównym problemem przy poszukiwaniu skutecznej ochrony na gruncie istniejącej regulacji będzie różnicowanie opierające się na kategoriach i danych nieuznanych za cechy chronione czy na zespołach tych cech. Dane i informacje w procesie przetwarzania umożliwiają generowanie dynamicznych profili jednostek, to zaś utrudnia identyfikację i ochronę pokrzywdzonych różnicowaniem. Dodatkowo z powodu zdolności do uczenia się zaawansowanych systemów opartych na analizie danych takie procesy decyzyjne będą wymykać się nadzorowi ludzkiemu, przez co instrumenty prawa posadowione na dotychczasowym rozumieniu dyskryminacji nie zadziałają.

Procesy różnicowania, zarówno dla społeczeństwa jako całości, jak i dla dotkniętej nim jednostki, nie będą prowadziły do widocznej stygmatyzacji czy segregacji, albowiem ofiara niesprawiedliwego procesu nie będzie wiedziała, że jej prawa kształtowane są mniej korzystnie. Nie będzie pojmowała katalogu cech prowadzących do zakwalifikowania jej do konkretnej grupy, a tak utworzonej społeczności nie będzie łączyła wspólna tożsamość. Dyskryminacja automatyczna nie będzie zatem wyzwalała społecznego sprzeciwu prowadzącego do szerszego buntu, analogicznego do opisywanych w rozdziale I przykładów historycznej walki o równość. Charakterystyka procesu różnicowania utrudnia więc nie tylko

[415] B. Mittelstadt *et. al.*, *op. cit*.

119:31050

identyfikację, ale również mitygację ryzyk, a w konsekwencji wpływa na upośledzenie dotychczasowych metod nadzoru i wykonywania władztwa.

Realizacja prawa do niepodlegania dyskryminującemu procesowi decyzyjnemu w toku stosowania ASD jest uwarunkowana przede wszystkim rozpoznaniem nieprawidłowości. Jak ukazuje przegląd orzecznictwa, z powodu nieprzejrzystości procesów i ich niejawności sądy zazwyczaj nie znajdowały podstaw do stosowania przepisów antydyskryminacyjnych. Wobec widocznej niesprawiedliwości posiłkowały się narzędziami z zakresu ochrony danych osobowych czy prawa do prywatności. Żadna z tych regulacji nie obejmowała jednak problemu kompleksowo. Egzekwowanie obowiązków przejrzystości ASD czy nadzoru ludzkiego może się zaś okazać pozorne wobec specyfiki głębokiego uczenia algorytmów. Wobec zwiększającej się skali problemu, na skutek m.in. intensywnych prac nad rozwojem sztucznej inteligencji, będącej zaawansowaną formą automatycznych stystemów decyzynych, należy w związku z tym przeanalizować kierunek i treść projektowanej regulacji dotyczącej zaawansowanych ASD, w tym SI.

W tym rozdziale zawarłam analizę prac legislacyjnych w Unii Europejskiej dotyczących uregulowania sztucznej inteligencji, będącej zaawansowaną technologicznie formą automatycznych systemów decyzyjnych, jak również starałam się ustalić adekwatność regulacji do opisanego problemu algorytmicznego różnicowania, w tym zagrożeń w postaci masowości, niejawności i nieprzejrzystości algorytmicznego różnicowania, ograniczonej możliwości zaobserwowania skutków rozstrzygnięć ASD w chwili różnicowania oraz niekoherentności różnicowania algorytmicznego z aktualnym zakresem i formami zakazanej dyskryminacji.

1. Droga do uregulowania sztucznej inteligencji

ASD w ograniczonym zakresie podlegały przepisom RODO i to ta regulacja – ze względu na jej refleksyjny charakter[416] – dała asumpt i wyznaczyła kierunki rozważań nad sposobem oraz mechanizmem ograniczania ryzyk związanych z automatyzacją. Innowacyjne podejście do ryzyka realizowane w ramach RODO zbiegło się przy tym z zauważalnym przyspieszeniem postępu technologicznego obejmującego rozwój zaawansowanych ASD, w tym systemów

[416] M. Pichlak, K. Gaczoł, *op. cit.*

120:68144

sztucznej inteligencji. To zaś doprowadziło do dyskusji prawnych, etycznych i praktycznych związanych z upowszechnieniem tychże systemów.

W związku z wyzwaniami dotyczącymi ochrony praw podstawowych na terenie Unii Europejskiej podjęto prace mające na celu ujednolicenie i wzmocnienie regulacji przy jednoczesnym wspieraniu innowacji w zakresie sztucznej inteligencji[417]. Unia dostrzegła, że systemy SI z dużym prawdopodobieństwem nie ominą żadnego obszaru życia, żadnej warstwy społecznej czy gałęzi gospodarki. Konieczna stała się zatem koordynacja działań na poziomie unijnym w celu urealnienia konkurencyjności państw członkowskich.

Początkowo prace legislacyjne i konsultacje w sprawie regulacji odbywały się w grupach eksperckich (zarówno na poziomie państwowym, jak i europejskim), aby następnie wkroczyć w etap tworzenia prawa miękkiego (ang. *soft law*) i projektów regulacji. Dokumentem prawnym otwierającym dyskusję co do kształtu i zakresu regulacji sztucznej inteligencji na gruncie prawa europejskiego była przyjęta 16.02.2017 r. rezolucja Parlamentu Europejskiego skierowana do Komisji z zaleceniami wypracowania rozwiązań (przepisów) w zakresie prawa cywilnego dotyczących robotyki[418]. Przyjęcie rezolucji było poprzedzone licznymi działaniami, które miały umożliwić rozwój gospodarki opartej na danych[419]. Następnie Komisja Europejska (dalej: KE) przyjęła skoordynowany plan[420] w sprawie rozwoju i wykorzystania sztucznej inteligencji opracowanej w Europie. Opierając się na pracach grupy niezależnych ekspertów, KE w grudniu 2018 r. opublikowała wstępną listę wskazówek dla rozwoju SI (m.in. zakaz identyfikacji użytkownika bez jego zgody, transparentność w stosowaniu SI, zachowanie kontroli człowieka nad wytworami techniki). Wynik tych prac

[417] Wobec braku wspólnych rządowych wysiłków na rzecz uregulowania SI w USA prywatne firmy zaczęły działać samodzielnie. W październiku 2017 r. DeepMind, wiodąca na świecie firma zajmująca się sztuczną inteligencją, przejęta przez Google w 2014 r., powołała nową radę ds. etyki, „aby pomóc technologom zastosować etykę w praktyce oraz pomóc społeczeństwu przewidzieć i ukierunkować wpływ AI, tak aby działała ona z korzyścią dla wszystkich" – A. Hern, *DeepMind Announces Ethics Group to Focus on Problems of AI*, „The Guardian" 4.10.2017, https://www.theguardian.com/technology/2017/oct/04/google-deepmind-ai-artificial-intelligence-ethics-group-problems (4.12.2021). Podobnie w 2016 r. siedem największych firm technologicznych – Amazon, Apple, DeepMind, Google, Facebook, IBM i Microsoft – podjęło próbę zbadania i sformułowania najlepszych praktyk w zakresie technologii sztucznej inteligencji.

[418] Rezolucja Parlamentu Europejskiego z 16.02.2017 r. zawierająca zalecenia dla Komisji w sprawie przepisów prawa cywilnego dotyczących robotyki, P8_TA(2017)0051, Dz.Urz. UE C 252 z 18.07.2018 r., s. 239–257.

[419] Rozporządzenie (UE) 2016/679; rozporządzenie (UE) 2018/1807; rozporządzenie (UE) 2019/881; dyrektywa (UE) 2019/1024; rozporządzenie 715/2007 zmienione rozporządzeniem 595/2009; dyrektywa 2015/2366 w sprawie usług płatniczych; dyrektywa 2019/944 w odniesieniu do energii elektrycznej; dyrektywa 2009/73/WE w odniesieniu do gazomierzy; rozporządzenie Komisji (UE) 2017/1485; rozporządzenie Komisji (UE) 2015/703; dyrektywa 2010/40/UE oraz dyrektywa (UE) 2019/770.

[420] Komunikat Komisji do Parlamentu Europejskiego, Rady Europejskiej, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów: *Sztuczna inteligencja dla Europy*, COM(2018) 237 final, https://eur-lex.europa.eu/legal-content/PL/TXT/?uri=COM%3A2018%3A237%3AFIN (dostęp: 4.12.2021).

121:84270

można poczytywać za postulaty wobec faktycznych możliwości i zakresu kontroli ludzkiej nad ASD.

W związku z tym Komisja Europejska formalnie powołała grupę ekspertów wysokiego szczebla, która w kwietniu 2019 r. opublikowała wytyczne w sprawie godnej zaufania sztucznej inteligencji[421]. Według ekspertów godna zaufania sztuczna inteligencja powinna mieć trzy cechy:

1) powinna być zgodna z prawem, tj. zapewniać poszanowanie wszystkich obowiązujących przepisów ustawowych i wykonawczych;
2) powinna być etyczna, zapewniając poszanowanie dla zasad i wartości etycznych, a także zgodność z nimi;
3) powinna być solidna zarówno z technicznego, jak i społecznego punktu widzenia, ponieważ systemy SI mogą wywołać niezamierzone szkody nawet wówczas, gdy korzysta się z nich w dobrej wierze.

Godna zaufania sztuczna inteligencja oznacza przy tym te same standardy nie tylko w odniesieniu do systemów SI, ale także do wszystkich procesów i podmiotów, które są objęte cyklem życia danego systemu.

W wydanych w 2019 r. konkluzjach do tego planu[422] Rada Unii Europejskiej podkreśliła m.in. znaczenie zapewnienia pełnego poszanowania praw europejskich obywateli poprzez wdrożenie wytycznych w zakresie etyki przy rozwijaniu i wykorzystywaniu sztucznej inteligencji w Unii Europejskiej i na szczeblu globalnym, czyniąc z aspektu etycznego w sztucznej inteligencji element dający europejskiemu przemysłowi przewagę konkurencyjną. Jednocześnie wezwała do zmiany, w stosownych przypadkach, istniejących odnośnych przepisów zgodnie z zasadami lepszego stanowienia prawa w celu zapewnienia, by były one dostosowane do nowych możliwości i wyzwań generowanych przez SI, w tym w zakresie takich kwestii jak bezpieczeństwo, prywatność, odpowiedzialność oraz decyzje i działania podejmowane bez udziału człowieka.

[421] Niezależna grupa ekspertów wysokiego szczebla ds. sztucznej inteligencji powołana przez Komisję Europejską w czerwcu 2018 r., *op. cit.*
[422] Rada Unii Europejskiej, *Sztuczna inteligencja, b) Konkluzje na temat skoordynowanego planu w sprawie sztucznej inteligencji – Przyjęcie*, 6177/19, 2019.

122:28420

Komisja opublikowała następnie komunikat[423] dotyczący budowania zaufania do sztucznej inteligencji ukierunkowanej na człowieka, w którym z zadowoleniem przyjęła siedem kluczowych wymogów określonych w wytycznych grupy ekspertów wysokiego szczebla:

1) przewodnia i nadzorcza rola człowieka;
2) techniczna solidność i bezpieczeństwo;
3) ochrona prywatności i zarządzanie danymi;
4) przejrzystość;
5) różnorodność, niedyskryminacja i sprawiedliwość;
6) dobrostan społeczny i środowiskowy;
7) odpowiedzialność.

Jedną z najistotniejszych wytycznych jest tutaj podejście ukierunkowane na człowieka, który jest i winien pozostać w centrum zainteresowania przyszłej legislacji. Oznacza to, że przy tworzeniu SI należy zagwarantować ludzkie wartości w sposobie opracowywania, wdrażania, wykorzystywania i monitorowania systemów SI poprzez zapewnienie poszanowania praw podstawowych, w tym tych zapisanych w traktatach i Karcie Praw Podstawowych Unii Europejskiej. SI ma być zatem również osadzona w poszanowaniu godności ludzkiej, w ramach której istocie ludzkiej przypisany jest wyjątkowy i niezbywalny status moralny.

Przełomowym dokumentem okazała się przyjęta przez Komisję 19.02.2020 r. Biała Księga w sprawie sztucznej inteligencji. Zawiera ona analizę wpływu SI, Internetu rzeczy i innych technologii cyfrowych na przepisy dotyczące bezpieczeństwa i odpowiedzialności. Dokument ten jest obszerny i odnosi się do wielu kluczowych aspektów bezpieczeństwa w kontekście stosowania technologii SI. Co ważne, wnioski formułowane w Białej Księdze zostały poprzedzone badaniami, m.in. ankietami skierowanymi do ok. 350 organizacji. W ich następstwie stwierdzono, że przepisy dotyczące przejrzystości, identyfikowalności i nadzoru przez człowieka nad SI nie zostały uwzględnione w obecnym prawodawstwie w odniesieniu do korzystania ze sztucznej inteligencji w wielu sektorach gospodarki. W Białej Księdze stwierdzono również, że zagrożenia dotyczące praw podstawowych mogą być wynikiem zarówno wadliwego opracowania systemów SI, jak i wykorzystywania danych do uczenia systemów bez korygowania uprzedzeń w nich zapisanych lub w taki sposób, że dane te będą prowadziły do niesprawiedliwych wyników analizy (np. system jest szkolony z

[423] Komunikat Komisji do Parlamentu Europejskiego, Rady, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów: *Budowanie zaufania do sztucznej inteligencji ukierunkowanej na człowieka*, COM(2019) 168.

123:12556

wykorzystaniem wyłącznie lub głównie danych pochodzących od konkretnych grup społecznych, co może prowadzić do niereprezentatywnych wyników). W związku z tym uznano, że dyskryminacja może obejmować istotne części społeczeństwa w sposób dotychczas niespotykany.

Dodatkowo Unia podkreśliła, że należy zwrócić szczególną uwagę na pokusę nieuzasadnionego, a przy tym masowego nadzoru i inwigilacji obywateli przez władzę publiczną. Wskazano również, że systemy SI mogą być wykorzystywane do znajdowania źródeł pochodzenia danych i deanonimizacji danych dotyczących osób, co uzasadnia konieczność zabezpieczenia również tzw. danych nieosobowych. Mając na względzie specyfikę SI, niemożliwe może się jednak okazać znalezienie przyczyny podejmowania błędnych decyzji przez ASD. Brak realnej kontroli wpływa zaś na trudności w wypracowaniu nowych zasad odpowiedzialności. W tym zakresie postanowiono poszukiwać odpowiednich metod ochrony jednostki poprzez:

1) zapewnienie egzekwowania przepisów (utworzenie zasad przypisywania i udowadniana odpowiedzialności za działania SI);
2) objęcie przepisami dotyczącymi bezpieczeństwa produktów również samego oprogramowania;
3) doprecyzowanie zasad ponoszenia odpowiedzialności za systemy SI nie tylko na etapie ich wprowadzenia do obrotu, ale też w toku ich aktualizacji czy udoskonalania na skutek uczenia maszynowego;
4) doprecyzowanie obowiązków poszczególnych podmiotów gospodarczych, szczególnie zasad ponoszenia odpowiedzialności w łańcuchu dostaw, kiedy to systemy SI będą dostarczane przez podmioty niebędące ich producentami;
5) budowanie bazy dowodowej na temat potencjalnych zagrożeń związanych z SI, zwłaszcza w kontekście cyberbezpieczeństwa i bezpieczeństwa osobistego[424].

Dostrzegając nieadekwatność istniejących narzędzi prawnych, Rada Unii Europejskiej w konkluzjach z 21.10.2020 r.[425] wezwała dodatkowo do sprostania takim wyzwaniom jak: efekt czarnej skrzynki, złożoność, stronniczość, pewna nieprzewidywalność oraz częściowo autonomiczne zachowanie niektórych systemów sztucznej inteligencji w celu zapewnienia ich

[424] Biała Księga w sprawie sztucznej inteligencji: *Europejskie podejście do doskonałości i zaufania*, COM(2020) 65.
[425] Rada Unii Europejskiej, Konkluzje prezydencji – Karta Praw Podstawowych w kontekście sztucznej inteligencji i przemian cyfrowych, 11481/20, 2020.

124:12957

zgodności z prawami podstawowymi i ułatwienia egzekwowania przepisów. Wobec nieprzejrzystości systemów nowe ramy regulacyjne, podobnie jak RODO, miały być oparte na analizie ryzyka, przy czym postuluje się, aby określenie wysokiego ryzyka było jasne i łatwe do zrozumienia oraz możliwe do zastosowania przez uczestników obrotu[426].

Jednocześnie UE proponuje, aby wysokie ryzyko istniało w przypadku spełnienia łącznie dwóch przesłanek, mianowicie zastosowania SI w sektorze, który może generować znaczące ryzyko, oraz w sposób zwiększający prawdopodobieństwo wystąpienia ryzyka[427]. Co do zasady regulacje byłyby bowiem aplikowane jedynie przy wykorzystaniu SI charakteryzującej się wysokim ryzykiem. Wyjątkami od tych zasad byłoby stosowanie systemów sztucznej inteligencji przy rekrutacji oraz w sposób mogący mieć wpływ na prawa pracowników (ochrona równości zatrudnienia) oraz użytkowanie SI do zdalnej identyfikacji biometrycznej czy innych inwazyjnych technologii nadzoru. Pod rozwagę poddano dodatkowe wyłączenie, które byłoby związane z ochroną konsumentów.

Unia Europejska przedstawiła również obowiązkowe wymogi prawne – zgodnie z wytycznymi grupy ekspertów wysokiego szczebla – dzielone na następujące grupy:

1) dane szkoleniowe;
2) przechowywanie danych i prowadzenie rejestrów oraz wymagane informacje;
3) solidność i dokładność;
4) sprawowanie nadzoru przez człowieka;
5) szczególne wymogi dotyczące określonych zastosowań sztucznej inteligencji, np. do celów zdalnej identyfikacji biometrycznej.

Nie dokonując szczegółowego omówienia tych grup, warto zaznaczyć, że Unia dostrzega konieczność przeciwdziałania dyskryminacji i niepodważania autonomii człowieka oraz ochrony prywatności i danych osobowych[428].

Na podstawie zebranych w Białej Księdze założeń Parlament Europejski podjął inicjatywę mającą na celu wezwanie Komisji Europejskiej do przedstawienia nowych ram prawnych

[426] Warto zwrócić uwagę na podejście niemieckie, w ramach którego wezwano do przygotowania wielostopniowego systemu regulacji, opartego na ocenie ryzyka – od braku regulacji SI nieszkodliwej po zupełny zakaz stosowania niebezpiecznych algorytmów.
[427] W Białej Księdze jako przykład podano użycie SI w sektorze medycznym, gdzie np. system umawiania wizyt sam w sobie nie będzie potencjalnym źródłem ryzyka, uzasadniającym interwencję legislacyjną.
[428] Łącznie z Białą Księgą został opublikowany komunikat Komisji do Parlamentu Europejskiego, Rady, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów: *Europejska strategia w zakresie danych*, COM(2020) 66.

125:10733

określających zasady etyczne i zobowiązania prawne, których należy przestrzegać przy opracowywaniu, wdrażaniu i wykorzystywaniu sztucznej inteligencji, robotyki i powiązanych technologii w UE, w tym oprogramowania, algorytmów i danych[429]. Stworzony dzięki temu dokument zawiera podsumowanie założeń i analiz zawartych dotychczas w Białej Księdze oraz poprzedzających ją komunikatach i opiniach na temat SI. Parlament przyjął przy tym, że wymiar prawny i etyczny należy zapisać w skutecznych, przyszłościowych i kompleksowych ramach regulacyjnych na szczeblu Unii, wspieranych przez właściwe organy krajowe, koordynowanych i wspieranych przez Komisję lub właściwe instytucje, organy i jednostki organizacyjne Unii, które mogą zostać wyznaczone w tym kontekście, regularnie wspomagane przez wspomniany ośrodek ekspercki, oraz należycie przestrzeganych i certyfikowanych w ramach rynku wewnętrznego.

Projekt stanowiska[430] dotyczącego sztucznej inteligencji, odnoszącego się m.in. do aspektów etycznych, przedstawiono 21.10.2020 r. Jego założenia miały zostać wpisane do Karty Praw Podstawowych, zawierającej podstawowe prawa obywateli Unii Europejskiej, w efekcie czego doszłoby do zaktualizowania treści Karty. Stanowisko nie zostało jednak przyjęte[431] z uwagi na sprzeciw Polski, co należy ocenić krytycznie. Odrzucenie przez Polskę wspólnej drogi do ujednolicenia rynku cyfrowego oraz wsparcia prac badawczych i naukowych w dziedzinie SI z dużym prawdopodobieństwem będzie negatywnie oddziaływać przede wszystkim na krajową gospodarkę. Oczywiste jest bowiem – bez względu na ocenę możliwości wdrożenia założeń prezentowanych obecnie w projektach Unii Europejskiej – że samodzielnie żaden z krajów Unii nie może realnie konkurować z potentatami na rynku systemów sztucznej inteligencji. Co więcej, KPP jest głównym dokumentem, do którego w zakresie ochrony praw i wolności jednostki odwołują się inne unijne akty prawne. Z tej perspektywy pożądanym i słusznym kierunkiem byłoby uregulowanie ram etycznej SI na tym szczeblu.

Oprócz głównego dokumentu, jakim była Biała Księga, w październiku 2020 r. Unia Europejska przyjęła zestaw rezolucji związanych ze sztuczną inteligencją, w tym w sprawie

[429] Rezolucja Parlamentu Europejskiego z 20.10.2020 r. zawierająca zalecenia dla Komisji w sprawie ram aspektów etycznych sztucznej inteligencji, robotyki i powiązanych z nimi technologii, 2020/2012(INL).
[430] Stanowisko unijne w sprawie sztucznej inteligencji (bez Polski) zob. *Presidency Conclusions, The Charter of Fundamental Rights in the Context of Artificial Intelligence and Digital Change*, https://polit-x.de/en/documents/4580568/ (dostęp: 4.12.2021).
[431] Polska sprzeciwiła się określeniu *gender equality* jako tłumaczeniu pojęcia „równouprawnienie".

126:29110

aspektów etycznych[432], odpowiedzialności[433] i praw autorskich[434]. Następnie w 2021 r. przyjęto rezolucje w sprawie sztucznej inteligencji w sprawach karnych[435], a także obowiązujące w sektorze edukacji i kultury oraz w sektorze audiowizualnym[436]. W rezolucji Parlamentu Europejskiego w sprawie ram aspektów etycznych sztucznej inteligencji, robotyki i powiązanych z nimi technologii zalecono zwłaszcza, aby Komisja zaproponowała działania ustawodawcze w celu wykorzystania możliwości i korzyści związanych ze sztuczną inteligencją, a jednocześnie w celu zapewnienia ochrony zasad etycznych. Rezolucja ta zawiera ponadto tekst wniosku ustawodawczego odnośnie do rozporządzenia w sprawie zasad etycznych dotyczących opracowywania, wdrażania i wykorzystywania sztucznej inteligencji, robotyki i powiązanych z nimi technologii.

W kolejnym roku – 3.05.2022 r. – Parlament Europejski przyjął rezolucję w sprawie sztucznej inteligencji w epoce cyfrowej[437], w której opisano ryzyko stronniczych decyzji prowadzących do dyskryminacji oraz łamania praw człowieka i konieczność refleksji nad kierunkiem regulacji. Potwierdzono również, że wszelkie formy normatywnej oceny punktowej obywateli przez organy publiczne prowadzą do utraty autonomii i prywatności, a także niosą z sobą ryzyko dyskryminacji i nie są zgodne z wartościami europejskimi. Na tym tle sformułowano postulaty i określono priorytety działania wobec inherentnych i masowych ryzyk związanych z różnicowaniem przez SI, przy czym postulaty te powielały dotychczasowe ustalenia, wypracowane m.in. w ramach Białej Księgi.

2. Akt w sprawie sztucznej inteligencji

[432] Rezolucja Parlamentu Europejskiego z 20.10.2020 r. w sprawie ram aspektów etycznych sztucznej inteligencji, robotyki i powiązanych z nimi technologii, 2020/2012(INL).
[433] Rezolucja Parlamentu Europejskiego z 20.10.2020 r. w sprawie systemu odpowiedzialności cywilnej za sztuczną inteligencję, 2020/2014(INL).
[434] Rezolucja Parlamentu Europejskiego z 20.10.2020 r. w sprawie praw własności intelektualnej w dziedzinie rozwoju technologii sztucznej inteligencji, 2020/2015(INI).
[435] Projekt sprawozdania Parlamentu Europejskiego na temat sztucznej inteligencji w prawie karnym i jej stosowania przez policję i organy wymiaru sprawiedliwości w sprawach karnych, 2020/2016(INI).
[436] Projekt sprawozdania Parlamentu Europejskiego w sprawie sztucznej inteligencji w sektorze edukacji i kultury oraz w sektorze audiowizualnym, 2020/2017(INI). W tym kontekście Komisja przyjęła *Plan działania w dziedzinie edukacji cyfrowej na lata 2021–2027. Nowe podejście do kształcenia i szkolenia w epoce cyfrowej*, w którym przewidziano opracowanie wytycznych etycznych dotyczących sztucznej inteligencji i wykorzystania danych w kształceniu; zob. komunikat Komisji do Parlamentu Europejskiego, Rady, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów, COM(2020) 624 final.
[437] Rezolucja Parlamentu Europejskiego z 3.05.2022 r. w sprawie sztucznej inteligencji w epoce cyfrowej: *Sztuczna inteligencja w epoce cyfrowej*, 2020/2266(INI).

127:88580

Komisja Europejska 21.04.2021 r. złożyła wniosek obejmujący pierwszy projekt rozporządzenia Parlamentu Europejskiego i Rady ustanawiającego zharmonizowane przepisy dotyczące sztucznej inteligencji (tzw. Akt w sprawie sztucznej inteligencji; ang. *Artificial Intelligence Act* – AIA) i zmieniającego niektóre akty ustawodawcze Unii[438]. Co istotne, rozporządzenie jako akt prawa Unii Europejskiej nie będzie wymagało transpozycji krajowych[439]. Nadto w zakresie regulacji obejmujących wprowadzenie do obrotu, oddawanie do użytku oraz wykorzystywanie systemów sztucznej inteligencji rozporządzenie wyłączy kompetencje legislacyjne państw członkowskich UE[440]. W sprawach bez elementu transgranicznego i niewynikających z innych aktów prawnych Unii decyzje regulacyjne państw członkowskich będą ograniczone do systemów sztucznej inteligencji opracowanych lub wykorzystywanych wyłącznie do celów wojskowych (art. 2 ust. 3 AIA)[441].

Proponowane rozporządzenie jest częścią pakietu wniosków Komisji, które należy rozumieć łącznie, w tym:

1) projektu aktu o usługach cyfrowych (z przepisami dotyczącymi dostępu do danych badawczych)[442];
2) projektu aktu o rynku cyfrowym (z przepisami dotyczącymi sprzętu, systemów operacyjnych i dystrybucji oprogramowania związanego z SI)[443];
3) projektu rozporządzenia w sprawie produktów maszynowych (zmieniającego dyrektywę maszynową w odniesieniu do systemów SI, zdrowia, bezpieczeństwa oraz maszyn)[444];
4) projektu dotyczącego zmiany w zakresie odpowiedzialności za produkt w odniesieniu do systemów SI[445];

[438] Projekt rozporządzenia Parlamentu Europejskiego i Rady…, COM(2021) 206 final, procedura 2021/0106(COD).
[439] Art. 288 zd. 2 TFUE stanowi, że rozporządzenie ma zasięg ogólny, wiąże w całości i jest bezpośrednio stosowane we wszystkich państwach członkowskich.
[440] Zob. wyrok TSUE z 18.02.1970 r. w sprawie *Bollmann*, skarga nr 40/60, pkt 4.
[441] *Analiza wybranych aspektów projektu Aktu w sprawie sztucznej inteligencji*, 2021, s. 14, https://www.gov.pl/attachment/fa12287b-09c6-47d1-b0d6-7286c53d6917 (dostęp: 4.12.2021).
[442] Wniosek: rozporządzenie Parlamentu Europejskiego i Rady w sprawie jednolitego rynku usług cyfrowych (akt o usługach cyfrowych) i zmieniające dyrektywę 2000/31/WE, COM(2020) 825 final.
[443] Wniosek: rozporządzenie Parlamentu Europejskiego i Rady w sprawie kontestowalnych i uczciwych rynków w sektorze cyfrowym (Akt o rynkach cyfrowych), COM(2020) 842 final.
[444] Wniosek: rozporządzenie Parlamentu Europejskiego i Rady w sprawie produktów maszynowych, COM(2021) 202 final.
[445] Komunikat Komisji do Parlamentu Europejskiego, Rady, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów: *Promowanie europejskiego podejścia do sztucznej inteligencji*, COM(2021) 205 final.

128:62309

5) projektu aktu w sprawie zarządzania danymi (w tym ram udostępniania danych)[446].

Prace legislacyjne nad projektem nadal trwają, a w ich toku doszło do wielokrotnych uzupełnień i zmiany pierwotnej wersji wniosku, w tym na skutek inicjatywy zarówno Prezydencji Rady, jak i Komisji. Kolejna wersja projektu (tzw. kompromisowa) została opublikowana 29.11.2021 r., a następna – 20.04.2022 r. Obecnie prace nad ostatecznym brzmieniem są przeprowadzane równolegle w Radzie UE i Parlamencie. Mimo wprowadzania już wielu zmian do regulacji na etapie konsultowania projektów i istotnego prawdopodobieństwa dalszych modyfikacji do celów niniejszej analizy przyjmuję, że kluczowe rozwiązania leżące u podstaw projektu nie ulegną zmianie. Obecne zmiany oscylują bowiem przede wszystkim wokół definicji i pojęć wykorzystywanych w AIA.

Niewątpliwie kroki podjęte na poziomie Unii Europejskiej są konieczne, a ukierunkowanie regulacji ASD, w tym najbardziej zaawansowanych algorytmów SI, na człowieka, wzmacnianie ochrony praw podstawowych oraz tworzenie bezpiecznych, wiarygodnych oraz realizujących zasady etyczne systemów sztucznej inteligencji – pożądane. Trwające zmiany polityczne mogą przy tym zwiększyć prawdopodobieństwo, że AIA stanie się akceptowalnym modelem globalnym[447]. Zawarta w dalszej części tego podrozdziału analiza będzie zatem dotyczyć rozwiązań powstałych w toku prac nad projektami i dyskutowanych w środowiskach prawniczych w kontekście problemu dyskryminacji. Jednocześnie zaznaczam, że omawiane założenia są przedmiotem dalszych prac i konsultacji, w związku z czym mogą zostać zastąpione lub uzupełnione innymi instytucjami i narzędziami prawnymi.

AIA opiera się na art. 114 Traktatu o funkcjonowaniu Unii Europejskiej[448], w związku z czym jego celem powinno być ograniczanie barier handlowych. Równocześnie jednak wprowadza on liczne gwarancje dla przestrzegania praw ustanowionych Kartą Praw Podstawowych. Struktura aktu jest przy tym czymś nowym z perspektywy ochrony praw podstawowych, co może mieć istotne konsekwencje dla swobody działania państw członkowskich[449]. W tym

[446] Wniosek: rozporządzenie Parlamentu Europejskiego i Rady w sprawie europejskiego zarządzania danymi (Akt w sprawie zarządzania danymi), COM(2020) 767 final.
[447] L. Edwards, *The EU AI Act: A Summary of Its Significance and Scope*, Expert Explainer 2022, s. 6, https://www.adalovelaceinstitute.org/wp-content/uploads/2022/04/Expert-explainer-The-EU-AI-Act-11-April-2022.pdf (dostęp: 4.12.2021).
[448] Art. 114. ust. 1 TFUE stanowi, że – z zastrzeżeniem, iż traktaty nie stanowią inaczej – do urzeczywistnienia celów określonych w art. 26 stosuje się następujące postanowienie: Parlament Europejski i Rada, stanowiąc zgodnie ze zwykłą procedurą ustawodawczą i po konsultacji z Komitetem Ekonomiczno-Społecznym, przyjmują środki dotyczące zbliżenia przepisów ustawowych, wykonawczych i administracyjnych państw członkowskich, które mają na celu ustanowienie i funkcjonowanie rynku wewnętrznego.
[449] *Analiza wybranych aspektów projektu*…, s. 16.

129:66912

kontekście należy wskazać, że deklarowanym celem rozporządzenia jest zapewnienie wysokiego stopnia ochrony praw podstawowych, szczególnie w odniesieniu do takich praw chronionych Kartą Praw Podstawowych jak: ochrona prawa do prywatności (art. 7 KPP), ochrona danych osobowych (art. 8 KPP), równe traktowanie (art. 21 KPP), a zwłaszcza zakaz dyskryminacji ze względu na płeć (art. 23 KPP). Jego celem jest także przeciwdziałanie tzw. efektowi mrożącemu, jeśli chodzi o wolność słowa (art. 11 KPP), czy też ochrona wolności zgromadzeń (art. 12 KPP). Podkreślono w nim również konieczność zabezpieczenia prawa do sądu (art. 47 KPP), domniemania niewinności (art. 48 KPP), co ma znaczenie w kontekście prawa karnego, oraz prawa do dobrej administracji (art. 41 KPP)[450].

Oparcie regulacji SI na art. 114 TFUE może zatem budzić wątpliwości. Przyjęcie rozporządzenia w związku z ograniczaniem barier handlowych powinno się bowiem koncentrować na perspektywie handlowej, a nie obywatelskiej. Ponadto na mocy tego przepisu nie przyznano prawodawcy unijnemu ogólnej kompetencji do regulowania rynku wewnętrznego[451]. Jak wskazał Trybunał Sprawiedliwości, wywodzenie takiej ogólnej kompetencji z art. 114 TFUE byłoby sprzeczne z przepisami definiującymi rynek wewnętrzny oraz z zasadą kompetencji powierzonych. Środki określone w tym przepisie muszą zaś mieć na celu poprawę warunków ustanowienia i funkcjonowania rynku wewnętrznego oraz powinny rzeczywiście zmierzać do jego urzeczywistnienia, przyczyniając się do zniesienia przeszkód w swobodnym przepływie towarów lub usług bądź do usunięcia zakłóceń konkurencji[452]. Struktura AIA w rzeczywistości nie wspiera zniesienia przeszkód, ponieważ regulacja stanowi realnie o nowych obowiązkach i zakazach. Co więcej, rozporządzenie obejmuje systemy sztucznej inteligencji wykorzystywane w obszarze zatrudnienia, zarządzania pracownikami i dostępu do samozatrudnienia, podczas gdy prawa i interesy pracowników mogą być określane przez Unię Europejską jedynie na podstawie art. 153 TFUE, który znacznie różni się od art. 114 TFUE co do zakresu kompetencji UE i procedur ustawodawczych, jakie należy zastosować[453].

Nie tylko idea ograniczenia barier handlowych przez wprowadzenie zestawu obowiązków i nakazów adresowanych do głównych aktorów procesu tworzenia, komercjalizacji i korzystania

[450] *Ibidem*, s. 14.
[451] A. Cieśliński, M. Szwarc, A. Zawidzka-Łojek, *Prawo rynku wewnętrznego. Funkcjonowanie Unii Europejskiej*, red. D. Kornobis-Romanowska, „System Prawa Unii Europejskiej", t. 7, Warszawa 2019, s. 168.
[452] Wyrok TSUE z 5.10.2000 r. w sprawie *Niemcy v. Parlament Europejski i Rada*, skarga nr C-376/98, pkt 83–84 oraz 95; wyrok TSUE z 10.12.2002 r. w sprawie *British American Tobacco (Investments) i Imperial Tobacco*, skarga nr C-491/01, pkt 60.
[453] *Analiza wybranych aspektów projektu*…, s. 16.

130:67696

z SI wydaje się trudna do uchwycenia. Akt prawny wydany na podstawie art. 114 TFUE nie może bowiem uzasadniać niepewnego i abstrakcyjnego ryzyka przeszkód w korzystaniu z podstawowych swobód lub występowania zakłóceń konkurencji, szczególnie w kontekście obecnego poziomu rozwoju SI i jej planowanej dalszej ewolucji. Opracowywana regulacja może z tych przyczyn formalnych, związanych z samą jej podstawą przyjęcia, być kwestionowana w przyszłości. Nadto jeżeli przyjąć, iż AIA jest źródłem zakazu dyskryminacji przez SI, może to dodatkowo – wobec omówionej w rozdziale III niniejszej pracy aktualnej regulacji europejskiego prawa antydyskryminacyjnego – mieć na celu modyfikację samego zakazu dyskryminacji. Jeżeli taki wymiar regulacji byłby ostatecznie realizowany, to pośrednio wpływałby na zakres zastosowania KPP.

Proponowany akt uzupełnia zatem istniejące regulacje unijne, nie stanowiąc jednak wyczerpującego źródła prawa dla systemów sztucznej inteligencji. Rozporządzenie w założeniu ma być przy tym spójne z postanowieniami Karty Praw Podstawowych Unii Europejskiej i obowiązującym prawem wtórnym Unii dotyczącym ochrony danych, ochrony konsumentów, niedyskryminacji i równouprawnienia płci. Oznacza to, że zakaz dyskryminacji istniejący w KPP będzie wyznaczał granice ochrony jednostki również w kontekście ASD objętych zakresem regulacji AIA. Deklarowana spójność w kontekście rozważań zawartych w rozdziale III nie jest jednak wystarczająca, przez co realny wpływ AIA na ochronę praw jednostki może być ograniczony. We wniosku wskazano co prawda, że – w założeniu – regulacja uzupełnia obowiązujące obecnie przepisy prawa Unii Europejskiej dotyczące niedyskryminacji o wymogi służące ograniczeniu do minimum ryzyka dyskryminacji automatycznej, szczególnie w zakresie projektowania i jakości zestawów danych wykorzystywanych do rozwoju systemów sztucznej inteligencji, poprzez ustanowienie obowiązków w zakresie testowania, zarządzania ryzykiem, dokumentacji i nadzoru ze strony człowieka w całym cyklu życia systemów sztucznej inteligencji.

Ustanowienie narzędzi do testowania czy oceny ryzyka nie jest jednak rozwiązaniem zadowalającym – w rozporządzeniu brakuje bowiem materialnoprawnej refleksji nad tym, czym jest dyskryminacja automatyczna w swoich specyficznych dla ASD formach i jak należy rozumieć wykluczenie algorytmiczne. W celu zapewnienia spójnego i wysokiego poziomu ochrony interesów publicznych w dziedzinie zdrowia, bezpieczeństwa i praw podstawowych preambuła projektowanego rozporządzenia przewiduje jedynie konieczność ustanowienia wspólnych standardów normatywnych dla wszystkich systemów sztucznej inteligencji wysokiego ryzyka. Standardy te zgodnie z treścią Aktu powinny być zgodne z KPP oraz być

131:76143

niedyskryminacyjne i zgodne z międzynarodowymi zobowiązaniami handlowymi Unii. AIA nie precyzuje jednak, które formy stronniczości algorytmicznej są zabronione w istniejących ramach prawnych oraz czy i w jakim stopniu należy je kwalifikować jako dyskryminację w rozumieniu KPP, a nawet przy użyciu jakich instrumentów stronniczość algorytmów powinna być łagodzona.

Nie każdy automatyczny system decyzyjny będzie przy tym objęty regulacją AIA. Pierwszy projekt AIA posługiwał się bardzo szeroką definicją systemów sztucznej inteligencji, rozumianych jako oprogramowanie opracowane przy użyciu co najmniej jednej spośród technik i podejść wymienionych w załączniku I, które może – dla danego zestawu celów określonych przez człowieka – generować wyniki, takie jak treści, przewidywania, zalecenia lub decyzje wpływające na środowiska, z którymi wchodzi w interakcję[454]. Pomimo zatem skupienia się AIA na systemach SI pierwotna regulacja realnie obejmowała bardzo szeroki zakres automatycznych systemów decyzyjnych. W toku prac legislacyjnych definicja SI uległa jednak zawężeniu z uwagi na pojawiające się głosy krytyczne[455]. W obecnym brzmieniu projektu proponowane jest, aby pod pojęciem systemów sztucznej inteligencji definiowane były systemy, które: 1. otrzymują dane lub wkład pochodzący od maszyny lub człowieka; 2. wnioskują, jak osiągnąć dany zestaw zdefiniowanych przez człowieka celów za pomocą uczenia się, rozumowania lub modelowania wdrażanych za pomocą technik i podejść wymienionych w załączniku I; 3. generują dane wyjściowe w postaci treści (tzw. *generative AI systems*), prognoz, zaleceń lub decyzji, które wpływają na środowiska, z którymi SI wchodzi w interakcje[456].

W załączniku I, do którego odwoływała się pierwotna definicja, zostały wskazane techniki i podejścia tworzące podstawę do tego, by zakwalifikować dany system jako system sztucznej inteligencji, takie jak: mechanizmy uczenia maszynowego, w tym uczenie nadzorowane, uczenie się maszyn bez nadzoru i uczenie przez wzmacnianie, z wykorzystaniem szerokiej gamy metod, w tym uczenia głębokiego, metod opartych na logice i wiedzy, w tym reprezentacji wiedzy, indukcyjnego programowania (logiczne), baz wiedzy, silników

[454] Art. 3 pkt 1 AIA.
[455] Grupa polskich ekspertów współpracujących z fundacją AI LAW TECH pod przewodnictwem Romana Biedy zaproponowała rozważanie zdefiniowania systemu SI w sposób bardziej opisowy, wskazujący na pewne atrybuty sztucznej inteligencji, sposoby czy cele działania takiego systemu, bez odwoływania się do konkretnych metod i technik; zob. *Analiza wybranych aspektów projektu*…, s. 10.
[456] Wniosek: rozporządzenie Parlamentu Europejskiego i Rady ustanawiające zharmonizowane przepisy dotyczące sztucznej inteligencji (Akt w Sprawie Sztucznej Inteligencji) oraz o zmianie niektórych unijnych aktów prawnych – Tekst kompromisowy prezydencji z 29.11.2021 r., 2021/0106(COD), nr 14278/21, art. 5.

132:67207

inferencyjnych i dedukcyjnych, rozumowania (symbolicznego) i systemów ekspertowych, podejścia statystycznego, estymacji bayesowskiej oraz metod wyszukiwania i optymalizacji.

Z perspektywy uregulowania zakazu dyskryminacji automatycznej, mogącej być efektem stosowania również mniej zaawansowanych systemów decyzyjnych, takie szerokie ujęcie przedmiotu regulacji było pożądane. Jednak z uwagi na to, że tak szeroka definicja wkraczała wprost w otoczenie regulacyjne m.in. programów komputerowych, ostatecznie zdecydowano się na jej ograniczenie. Definicja systemu SI nadal jest szeroka[457] – rozporządzenie nie ma zastosowania jedynie do systemów SI opracowanych lub wykorzystywanych wyłącznie do celów wojskowych lub bezpieczeństwa narodowego, jak również do prowadzenia badań naukowych i rozwojowych.

Uwaga dotycząca terminologii opracowanej do celów AIA ma tymczasem szersze odniesienie, jako że nie tylko pojęcie SI budziło kontrowersje. Uważa się bowiem, że terminy, którymi posłużono się w regulacji, są na tyle nieprecyzyjne, iż stosowanie rozporządzenia będzie niezwykle trudne[458]. Z tej też przyczyny ostateczny kształt art. 3 rozporządzenia, zawierającego słowniczek, może ulegać dalszym zmianom.

Regulacja zgodnie z jej pierwotnym projektem miała mieć zasięg eksterytorialny, podmiotowo obejmujący:

1) dostawców wprowadzających do obrotu lub oddających do użytku systemy sztucznej inteligencji w Unii, niezależnie od tego, czy dostawcy ci mają siedzibę w UE, czy w państwie trzecim;
2) użytkowników systemów sztucznej inteligencji, którzy znajdują się w Unii;
3) dostawców i użytkowników systemów sztucznej inteligencji, którzy znajdują się w państwie trzecim, jeżeli wyniki działania systemu są wykorzystywane w UE.

W toku prac legislacyjnych nad projektem AIA również krąg obowiązanych się zmienił – Parlament Europejski zaproponował bowiem zmianę kategorii dostawców na operatorów. Operatorem w myśl przepisów jest nie tylko dostawca, ale też upoważniony przedstawiciel, importer i dystrybutor. Zmianę tę należy ocenić pozytywnie, gdyż umożliwi ona objęcie regulacją wszystkich podmiotów wpływających na proces decyzyjny z wykorzystaniem SI.

[457] *Analiza wybranych aspektów projektu…*, s. 22.
[458] Zróżnicowana wykładnia pojęć, takich jak np. „istotne zniekształcenie zachowania", „szkoda psychiczna" czy „szkoda fizyczna", zagraża realizacji tego celu i może spowodować, że pewne modele biznesowe w niektórych państwach członkowskich będą dozwolone, a w innych nie; zob. *ibidem*, s. 15.

133:10036

Wobec tego akt znajdzie zastosowanie, nawet jeżeli algorytm będzie trenowany i rozwijany w państwie trzecim na zlecenie operatora mającego siedzibę w Unii albo przez operatora mającego siedzibę w państwie trzecim, w zakresie, w jakim wyniki działania takiego systemu będą wykorzystywane w Unii.

W rozporządzeniu zastosowano podejście oparte na analizie ryzyka, wprowadzając rozróżnienie między: 1. niedopuszczalnym ryzykiem, 2. wysokim ryzykiem oraz 3. niskim lub minimalnym ryzykiem. AIA nie definiuje jednak pojęcia ryzyka ani systemu zarządzania ryzykiem[459]. Konstrukcja oparcia systemu na kwalifikowaniu ryzyka przyjęta w AIA jest przy tym oparta na rozwiązaniach Nowych Ram Prawnych[460] (ang. *New Legislative Framework*; dalej: NLF) i jest z nimi powiązana[461]. Do polskiego porządku prawnego zasady NLF inkorporuje m.in. ustawa o systemach oceny zgodności i nadzoru rynku[462].

W zależności od zakwalifikowania systemu do poszczególnej grupy ryzyka rozporządzenie wprowadza odmienne regulacje prawne dla dostawców systemów, które mogą zupełnie zakazywać stosowania omawianych systemów, ograniczać ich stosowanie poprzez wprowadzenie dodatkowych obowiązków i procedur, a także dopuszczać wprowadzanie i korzystanie z systemów. Od 2022 r. kilka organów ustanawiających standardy opracowuje dobrowolne ramy zarządzania ryzykiem związanym z SI[463].

[459] Proponuję przyjąć tutaj ISO/IEC 51:2014: Aspekty bezpieczeństwa – wytyczne dotyczące ich włączenia do norm, zgodnie z którymi ryzyko można rozumieć jako implikację prawdopodobieństwa wystąpienia szkody i jej dotkliwości.

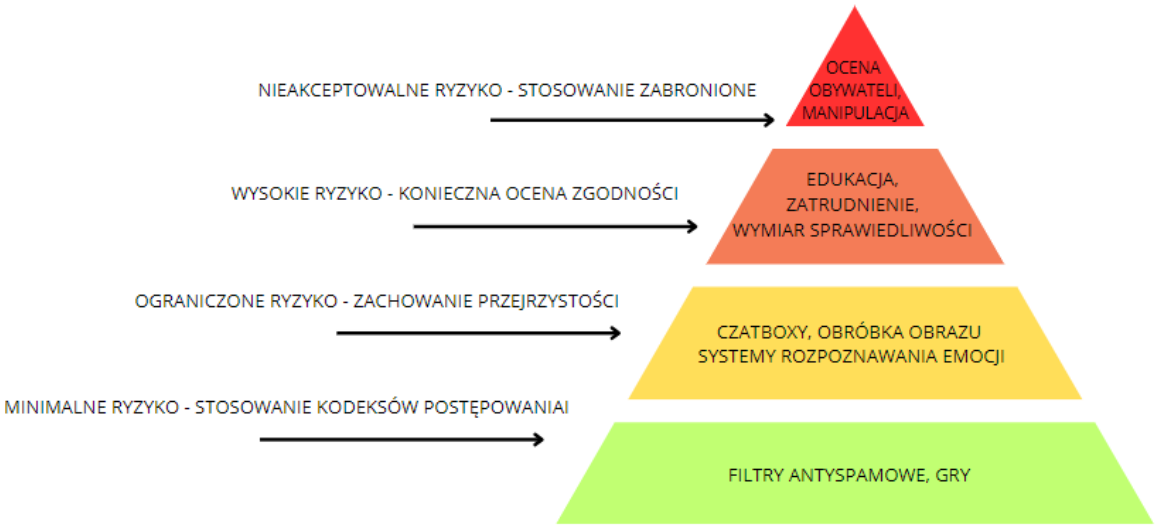
[460] W ramach systemów NLF (zob. https://ec.europa.eu/growth/single-market/goods/new-legislative-framework_pl, dostęp: 4.12.2021) producent musi przeprowadzać kontrole przed wprowadzeniem produktów na rynek, mające na celu ustalenie bezpieczeństwa i wydajności produktów poprzez ocenę zgodności z pewnymi zasadniczymi wymaganiami określonymi w przepisach prawa. Następnie producenci oznaczają produkty zgodne z wymogami znakiem CE (oznakowane produkty korzystają z unijnej swobody przepływu). Podejście NLF opiera się na przekonaniu, że producent, dysponując szczegółową wiedzą na temat projektu i procesu produkcji, jest w stanie najlepiej przeprowadzić pełną procedurę oceny zgodności. Por. M. Veale, F. Zuiderveen Borgesius, *Demystifying the Draft EU Artificial Intelligence Act. Analysing the Good, the Bad, and the Unclear Elements of the Proposed Approach*, „Computer Law Review International" 4, 2021, s. 97–112.

[461] Nowe podejście zostało wprowadzone w 1985 r. do regulacji niektórych produktów, takich jak dźwigi, wyroby medyczne, środki ochrony indywidualnej i zabawki; zob. *ibidem*.

[462] Ustawa o systemach oceny zgodności i nadzoru rynku z 13.04.2016 r., Dz.U. poz. 542, tekst jedn. Dz.U. z 2022 r. poz. 5.

[463] Najbardziej godne uwagi są NIST AI Risk Management Framework i ISO/IEC 23894; zob. J. Schuett, *Risk Management in the Artificial Intelligence Act*, „European Journal of Risk Regulation" 2023, s. 1–19.

134:32072



Rys. 4. Schemat podejścia opartego na ryzyku

Źródło: opracowanie własne na podstawie L. Edwards, *op. cit.*

Warto nadmienić, że piramida zaproponowana w toku prac nad AIA jest zbliżona do podejścia opartego na ryzyku prezentowanego wcześniej przez Niemcy.

Zgodnie z projektem rozporządzenia AIA systemy SI powodujące niedopuszczalne ryzyko objęte są zakazem wprowadzania do obrotu[464]. Pierwotnie były to praktyki, które wykazują znaczny potencjał manipulowania ludźmi, oparte są na technikach podprogowych działających na podświadomość ludzi albo wykorzystujących słabości określonych grup, takich jak dzieci lub osoby z niepełnosprawnościami, aby istotnie wypaczyć ich zachowania w sposób, który może spowodować u nich bądź u innej osoby szkodę psychiczną lub fizyczną.

Rozumienie manipulacji prezentowane w toku prac nad wnioskiem stało się jednak przedmiotem ostrej krytyki. Zarzucano m.in., że taki kształt regulacji doprowadzi do tego, że wnioskując z przeciwieństwa, manipulacyjne systemy SI staną się dozwolone, o ile jest mało prawdopodobne, aby wyrządziły jednostce (a nie zbiorowości) szkodę[465]. Zgodnie z głosami krytyki w ramach konsultacji przepisy dotyczące manipulacji zostały nieco wzmocnione poprzez osłabienie znaczenia zamiaru działania i uwzględnienie uzasadnionego prawdopodobieństwa szkody. Nadal jednak definicja systemów niedopuszczalnego ryzyka ma charakter nieostateczny. Co więcej, zawiera ona zestaw niejasnych, niejednoznacznych pojęć,

[464] Art. 5 AIA.
[465] M. Veale, F. Zuiderveen Borgesius, *op. cit.*

135:66761

które powodują daleko idącą niepewność prawną. Jak wskazuje się w dyskusji wokół projektu regulacji, może to mieć podwójnie negatywny efekt – po pierwsze, może spowodować zamrożenie i ograniczanie innowacji; po drugie, niejednoznaczna redakcja przepisów może prowadzić do prób omijania zakazu[466].

Druga grupa zakazów dotyczy tzw. scoringu społecznego. W pkt 17 preambuły rozporządzenia wyjaśniono, że systemy sztucznej inteligencji, które umożliwiają prowadzenie punktowania zachowań społecznych osób fizycznych do celów ogólnych, mogą prowadzić do dyskryminacyjnych wyników i wykluczenia pewnych grup. Mogą one bowiem naruszać prawo do godności i niedyskryminacji oraz takich wartości jak równość i sprawiedliwość. Takie systemy sztucznej inteligencji oceniają wiarygodność osób fizycznych na podstawie ich zachowań społecznych w wielu kontekstach lub na podstawie zespołów znanych czy też wywnioskowanych danych i informacji. Kategoria dobrana przez takie systemy sztucznej inteligencji może prowadzić do krzywdzącego lub niekorzystnego traktowania osób fizycznych lub ich grup w kontekstach społecznych, które nie były celem pierwotnego przetwarzania danych. Co więcej, naruszenie praw i wolności w toku różnicowania może być nieproporcjonalne lub nieuzasadnione w stosunku do znaczenia informacji. Tego rodzaju ryzyka zrealizowały się w przypadku SyRI czy polskiego ASD do profilowania bezrobotnych, jako że w wypadku SyRI zakwalifikowanie jednostki przez system do grupy ryzyka powodowało konieczność wszczęcia postępowania przez organy państwowe.

W pierwotnym założeniu zakaz miał mieć charakter zupełny, jednak zgodnie z obecnie proponowanym art. 5 lit. c) rozporządzenia zakaz uzyskał charakter względny. Oznacza to, że *scoring* jest zakazany w kontekście: 1. krzywdzącego lub niekorzystnego traktowania niektórych osób fizycznych lub całych ich grup w kontekstach społecznych, które nie są związane z kontekstami, w których pierwotnie wygenerowano lub zgromadzono dane; bądź 2. krzywdzącego lub niekorzystnego traktowania niektórych osób fizycznych lub całych ich grup, które jest nieuzasadnione albo nieproporcjonalne do ich zachowania społecznego lub jego wagi. Tego rodzaju zapis może jednak prowadzić do przekonania, że zakaz dyskryminacji rozumiany jako niedozwolone różnicowanie ze względu na chronione cechy będzie uzupełniany indywidualną oceną skutku różnicowania przez ASD w kontekście ewentualnego krzywdzącego lub niekorzystnego traktowania niektórych osób lub grup. Zapis ten jest niestety nieprecyzyjny i nie pozwala na jasne określenie zakresu zakazu.

[466] *Analiza wybranych aspektów projektu*…, s. 10.

136:97797

Krytyka takiego podejścia została zaprezentowana również przez polskich specjalistów z dziedziny sztucznej inteligencji, i to aż na czterech poziomach[467]. Po pierwsze, stwierdzono, że wprowadzenie jedynie względnego zakazu budzi poważne zastrzeżenia, zarówno co do koherencji całego AIA, jak i poszanowania praw podstawowych. Po drugie, wskazano, że nie jest jasne, kto i w jaki sposób ma w praktyce dokonywać oceny systemów sztucznej inteligencji stosujących *scoring* społeczny i jak ma być wdrażany taki zakaz. Po trzecie, podkreślono, że nie wiadomo, kto ma wykazać wystąpienie zakazanego skutku, by zaktualizował się zakaz zawarty w rozporządzeniu (kwestia ciężaru dowodu). Po czwarte, podniesiono, że zakres podmiotowy powoduje trudność w ustaleniu, do kogo w praktyce skierowany jest zakaz – nie zostało bowiem m.in. sprecyzowane, o jakiej natury związki/relacje między organami chodzi. Dyskusja wokół zakazu scoringu nadal trwa, w związku z czym wskazane wątpliwości mogą mieć charakter nieostateczny.

Rozporządzenie zawiera ponadto obszerną listę systemów SI klasyfikowanych jako „wysokiego ryzyka" (załącznik III)[468]. Wskazano przy tym, że przy kategoryzowaniu systemu sztucznej inteligencji do grupy wysokiego ryzyka[469] zasadnicze znaczenie ma skala szkodliwego wpływu wywieranego przez system SI na prawa podstawowe chronione na mocy KPP. Zgodnie z art. 6 ust. 1 projektu AIA jako systemy wysokiego ryzyka definiowane są systemy spełniające oba następujące warunki: 1. system sztucznej inteligencji jest przeznaczony do wykorzystywania jako związany z bezpieczeństwem element produktu objętego unijnym prawodawstwem harmonizacyjnym wymienionym w załączniku II[470] lub

[467] *Ibidem*, s. 28–29.

[468] Do praw tych należą: prawo do godności człowieka, poszanowanie życia prywatnego i rodzinnego, ochrona danych osobowych, wolność wypowiedzi i informacji, wolność zgromadzania i stowarzyszania się oraz niedyskryminacja, ochrona konsumentów, prawa pracownicze, prawa osób niepełnosprawnych, prawo do skutecznego środka prawnego i dostępu do bezstronnego sądu, prawo do obrony i domniemania niewinności, prawo do dobrej administracji; zob. art. 6–7 AIA.

[469] Klasyfikacja systemu sztucznej inteligencji jako systemu wysokiego ryzyka na podstawie rozporządzenia nie powinna koniecznie oznaczać, że produkt, którego związanym z bezpieczeństwem elementem jest system sztucznej inteligencji, lub sam system sztucznej inteligencji jako produkt uznaje się za produkt „wysokiego ryzyka" zgodnie z kryteriami ustanowionymi w stosownym unijnym prawodawstwie harmonizacyjnym, które ma zastosowanie do tego produktu.

[470] Systemy sztucznej inteligencji wysokiego ryzyka zgodnie z art. 6 ust. 2 AIA to systemy sztucznej inteligencji wymienione w którymkolwiek z następujących obszarów:

1. identyfikacja i kategoryzacja biometryczna osób fizycznych:
a) systemy sztucznej inteligencji przeznaczone do stosowania w celu zdalnej identyfikacji biometrycznej osób fizycznych w czasie rzeczywistym i *post factum*;
2. zarządzanie infrastrukturą krytyczną i jej eksploatacja:
a) systemy sztucznej inteligencji przeznaczone do stosowania jako związane z bezpieczeństwem elementy procesów zarządzania i obsługi ruchu drogowego oraz zaopatrzenia w wodę, gaz, ciepło i energię elektryczną;
3. kształcenie i szkolenie zawodowe:

137:69122

a) systemy sztucznej inteligencji przeznaczone do stosowania w celu podejmowania decyzji o dostępie do instytucji edukacyjnych i instytucji szkolenia zawodowego lub nadawania osobom przydziału do tych instytucji;
b) systemy sztucznej inteligencji przeznaczone do stosowania w celu oceny uczniów w instytucjach edukacyjnych i instytucjach szkolenia zawodowego oraz do oceny uczestników egzaminów powszechnie wymaganych w celu przyjęcia do instytucji edukacyjnych;

4. zatrudnienie, zarządzanie pracownikami i dostęp do samozatrudnienia:

a) systemy sztucznej inteligencji przeznaczone do wykorzystania w celu rekrutacji lub wyboru osób fizycznych, szczególnie w przypadku informowania o wakatach, selekcji lub filtrowania podań o pracę, oceny kandydatów w trakcie rozmów kwalifikacyjnych lub testów;
b) systemy sztucznej inteligencji przeznaczone do wykorzystania w celu podejmowania decyzji o awansie i rozwiązaniu stosunku pracy, przydzielania zadań oraz monitorowania i oceny wydajności czy zachowania osób pozostających w takich stosunkach;

5. dostęp do podstawowych usług prywatnych oraz usług i świadczeń publicznych, a także korzystanie z nich:

a) systemy sztucznej inteligencji przeznaczone do wykorzystania przez organy publiczne lub w imieniu organów publicznych w celu oceny kwalifikowalności osób fizycznych do świadczeń i usług publicznych, jak również w celu przyznawania, ograniczania, unieważniania lub żądania zwrotu takich świadczeń i usług;
b) systemy sztucznej inteligencji przeznaczone do wykorzystania w celu oceny zdolności kredytowej osób fizycznych lub ustalenia ich punktowej oceny kredytowej, z wyjątkiem systemów sztucznej inteligencji oddawanych do użytku przez drobnych dostawców na ich własny użytek;
c) systemy sztucznej inteligencji przeznaczone do wykorzystania w celu wysyłania lub ustalania priorytetów w wysyłaniu służb ratunkowych w sytuacjach kryzysowych, w tym straży pożarnej i pomocy medycznej;

6. ściganie przestępstw:

a) systemy sztucznej inteligencji przeznaczone do wykorzystania przez organy ścigania w celu przeprowadzania indywidualnych ocen ryzyka w odniesieniu do osób fizycznych, aby ocenić ryzyko popełnienia lub ponownego popełnienia przestępstwa przez osobę fizyczną bądź ocenić ryzyko, na jakie narażone są potencjalne ofiary przestępstw;
b) systemy sztucznej inteligencji przeznaczone do wykorzystania przez organy ścigania jako poligrafy i podobne narzędzia lub w celu wykrywania stanu emocjonalnego osoby fizycznej;
c) systemy sztucznej inteligencji przeznaczone do wykorzystania przez organy ścigania w celu wykrywania treści stworzonych z wykorzystaniem technologii *deepfake*, o których mowa w art. 52 ust. 3 dyrektywy (UE) 2016/680;
d) systemy sztucznej inteligencji przeznaczone do wykorzystania przez organy ścigania w celu oceny wiarygodności dowodów w toku ścigania przestępstw lub prowadzenia dochodzeń w ich sprawie;
e) systemy sztucznej inteligencji przeznaczone do wykorzystania przez organy ścigania w celu przewidywania wystąpienia lub ponownego wystąpienia rzeczywistego lub potencjalnego przestępstwa na podstawie profilowania osób fizycznych, o którym mowa w art. 3 pkt 4 dyrektywy (UE) 2016/680, lub w celu oceny cech osobowości i charakterystyki lub wcześniejszego zachowania przestępnego osób fizycznych lub grup;
f) systemy sztucznej inteligencji przeznaczone do wykorzystania przez organy ścigania w celu profilowania osób fizycznych, o którym mowa w art. 3 pkt 4 dyrektywy (UE) 2016/680, w toku wykrywania i ścigania przestępstw lub prowadzenia dochodzeń w ich sprawie;
g) systemy sztucznej inteligencji przeznaczone do wykorzystania do analizy przestępczości osób fizycznych, umożliwiające organom ścigania przeszukiwanie złożonych, powiązanych i niepowiązanych dużych zbiorów danych dostępnych w różnych źródłach danych lub w różnych formatach danych w celu zidentyfikowania nieznanych wzorców lub odkrycia ukrytych zależności między danymi;

7. zarządzanie migracją, azylem i kontrolą graniczną:

a) systemy sztucznej inteligencji przeznaczone do wykorzystania przez właściwe organy publiczne jako poligrafy i podobne narzędzia lub w celu wykrywania stanu emocjonalnego osoby fizycznej;
b) systemy sztucznej inteligencji przeznaczone do wykorzystania przez właściwe organy publiczne w celu oceny ryzyka, w tym zagrożenia dla bezpieczeństwa, ryzyka imigracji nieuregulowanej lub zagrożeń dla zdrowia stwarzanych przez osobę fizyczną, która zamierza wjechać lub wjechała na terytorium państwa członkowskiego;
c) systemy sztucznej inteligencji przeznaczone do wykorzystania przez właściwe organy publiczne w celu weryfikacji autentyczności dokumentów podróży i dokumentów uzupełniających osób fizycznych oraz wykrywania dokumentów nieautentycznych poprzez sprawdzanie ich zabezpieczeń;

sam jest takim produktem; 2. produkt związany z bezpieczeństwem jest elementem systemu sztucznej inteligencji lub sam system sztucznej inteligencji jest produktem. Na podstawie unijnego prawodawstwa harmonizacyjnego wymienionego w załączniku II podlegają one wówczas ocenie zgodności przeprowadzanej przez osobę trzecią w celu wprowadzenia tego produktu do obrotu lub oddania go do użytku.

Aktualnie projekt zakłada, że systemami wysokiego ryzyka są te, których rezultaty wywołują natychmiastowe skutki w odniesieniu do zamierzonego celu, bez potrzeby ich walidacji przez człowieka, jak również systemy, których rezultaty składają się z informacji będącej jedyną podstawą do podjęcia działań lub decyzji bądź których charakter nie jest wyłącznie akcesoryjny w odniesieniu do działania albo decyzji, które mają być podjęte przez człowieka, i z tych względów mogą stwarzać znaczne zagrożenie dla zdrowia, bezpieczeństwa lub praw podstawowych jednostek. Co więcej, za systemy wysokiego ryzyka uznaje się też systemy sztucznej inteligencji, o których mowa w załączniku III do rozporządzenia.

Tak rozproszona definicja systemów wysokiego ryzyka jest niejasna i trudna do zastosowania. System wysokiego ryzyka mógłby zostać określony w sposób bardziej opisowy zgodnie z zasadą neutralności technologicznej oraz podejściem opartym na ryzyku przy jednoczesnym pozostawieniu treści załącznika III[471].

Z perspektywy realizacji postulatu zakazu dyskryminacji istotne znaczenie mają następujące systemy zaliczane w projekcie do grupy wysokiego ryzyka:

1) systemy sztucznej inteligencji wykorzystywane w obszarze kształcenia lub szkolenia zawodowego, szczególnie przy podejmowaniu decyzji o dostępie do instytucji kształcenia i szkolenia zawodowego lub nadawaniu osobom przydziału do tych instytucji bądź do oceniania osób na testach w ramach kształcenia[472];
2) systemy sztucznej inteligencji wykorzystywane w obszarze zatrudnienia, zarządzania pracownikami i dostępu do samozatrudnienia, zwłaszcza do rekrutacji i wyboru kandydatów, do podejmowania decyzji o awansie i rozwiązaniu stosunku pracy oraz do

d) systemy sztucznej inteligencji, które mają służyć właściwym organom publicznym pomocą przy rozpatrywaniu wniosków o udzielenie azylu, wydanie wizy i dokumentów pobytowych oraz związanych z nimi skarg w odniesieniu do kwalifikowalności osób fizycznych ubiegających się o przyznanie określonego statusu;
8. sprawowanie wymiaru sprawiedliwości i procesy demokratyczne:
a) systemy sztucznej inteligencji, które mają służyć organowi sądowemu pomocą w badaniu i interpretacji stanu faktycznego i przepisów prawa oraz w stosowaniu prawa do konkretnego stanu faktycznego.

[471] *Analiza wybranych aspektów projektu*…, s. 10–11.
[472] Pkt 35 AIA.

139:55554

przydzielania zadań, monitorowania lub oceny osób pozostających w umownych stosunkach pracy[473];

3) systemy sztucznej inteligencji zapewniające dostęp do niektórych podstawowych usług i świadczeń prywatnych i publicznych niezbędnych ludziom do pełnego uczestnictwa w życiu społecznym lub do poprawy poziomu życia oraz korzystanie z tych usług i świadczeń (w tym systemy sztucznej inteligencji wykorzystywane do przeprowadzania punktowej oceny kredytowej lub oceny zdolności kredytowej)[474];
4) systemy sztucznej inteligencji wykorzystywane przez organy ścigania (w tym niektóre systemy SI przeznaczone do stosowania w kontekście egzekwowania prawa)[475];
5) systemy sztucznej inteligencji użytkowane w zarządzaniu migracją, azylem i kontrolą graniczną[476].

Jednocześnie rozporządzenie wskazuje, że również niektóre systemy sztucznej inteligencji opracowane na potrzeby sprawowania wymiaru sprawiedliwości i procesów demokratycznych należy sklasyfikować jako systemy wysokiego ryzyka, biorąc pod uwagę ich potencjalnie istotny wpływ na demokrację, praworządność, wolności osobiste, a także prawo do skutecznego środka odwoławczego i rzetelnego procesu sądowego[477].

Mimo że wyliczenie systemów wysokiego ryzyka ma charakter zamknięty, definicja może być rozszerzana co 24 miesiące na skutek przyjmowania przez Komisję aktów delegowanych w celu aktualizacji wykazu. Uprawnienie to z jednej strony jest krytykowane przez uczestników obrotu gospodarczego jako zbyt daleko idące i uderzające w pewność prawa oraz obrotu gospodarczego, z drugiej zaś oceniane jest pozytywnie jako narzędzie elastyczne, umożliwiające reakcję na postęp i zmieniające się otoczenie technologiczne. Zgodnie z art. 7 ust. 2 rozporządzenia Komisja, realizując to uprawnienie, ocenia, czy system sztucznej inteligencji stwarza ryzyko szkody dla zdrowia i bezpieczeństwa lub ryzyko niekorzystnego wpływu na prawa podstawowe, które jest równoważne z ryzykiem szkody stwarzanym przez systemy sztucznej inteligencji wysokiego ryzyka wymienione w załączniku III lub jest od niego większe. Przy czym taka analiza musi uwzględniać następujące kryteria:

1) przeznaczenie systemu sztucznej inteligencji;
2) zakres, w jakim system SI był wykorzystywany lub może być wykorzystywany;

[473] Pkt 36 AIA.
[474] Pkt 37 AIA.
[475] Pkt 38 AIA.
[476] Pkt 39 AIA.
[477] Pkt 40 AIA.

140:11490

3) zakres, w jakim zastosowanie systemu sztucznej inteligencji spowodowało już szkodę dla zdrowia i bezpieczeństwa lub miało niekorzystny wpływ na prawa podstawowe bądź wzbudziło istotne obawy co do możliwości wystąpienia takiej szkody lub niekorzystnego wpływu, czego potwierdzeniem są zgłoszenia lub udokumentowane zarzuty przedłożone właściwym organom krajowym;
4) potencjalny zakres takiej szkody lub takiego niekorzystnego wpływu, szczególnie pod względem ich nasilenia i możliwości oddziaływania na wiele osób;
5) zakres, w jakim osoby potencjalnie poszkodowane lub dotknięte niekorzystnym wpływem są zależne od wyniku działania systemu sztucznej inteligencji, zwłaszcza ze względu na fakt, że z przyczyn praktycznych lub prawnych nie jest zasadne zrezygnowanie z tego wyniku;
6) zakres, w jakim osoby potencjalnie poszkodowane lub dotknięte niekorzystnym wpływem znajdują się w słabszym położeniu względem innego użytkownika systemu sztucznej inteligencji, szczególnie z powodu nierównego układu sił, wiedzy, sytuacji gospodarczej lub społecznej albo wieku;
7) zakres, w jakim uzyskany za pomocą systemu sztucznej inteligencji wynik jest łatwo odwracalny, przy czym wyników działania systemu mających wpływ na zdrowie lub bezpieczeństwo osób za takie się nie uznaje.

Ostatnim kryterium kontroli Komisji zgodnie z treścią rozporządzenia jest zakres, w jakim obowiązujące przepisy Unii przewidują: 1. skuteczne środki dochodzenia roszczeń w związku z zagrożeniami stwarzanymi przez system sztucznej inteligencji, z wyłączeniem roszczeń o odszkodowanie; 2. skuteczne środki zapobiegania temu ryzyku lub jego znacznego zminimalizowania.

We wniosku dotyczącym rozporządzenia wskazano, że klasyfikacja systemu sztucznej inteligencji jako systemu wysokiego ryzyka opiera się na przeznaczeniu systemu SI zgodnie z obowiązującymi przepisami dotyczącymi bezpieczeństwa produktów. Dlatego też klasyfikacja jako system sztucznej inteligencji wysokiego ryzyka zależy nie tylko od funkcji pełnionej przez system SI, ale także od konkretnego celu i trybu jego wykorzystania. Powoduje to, że niektóre ASD, nawet jeżeli w świetle projektu rozporządzenia zostaną zakwalifikowane jako systemy SI, zasadniczo nie będą podlegały specyficznym wymaganiom zarezerwowanym głównie dla systemów wysokiego ryzyka, a to ze względu na wykorzystywane przez nie metody lub

141:85590

techniki[478]. W tym zakresie ponownie można zaobserwować ograniczony zakres zastosowania opracowywanej regulacji.

AIA przewiduje przy tym, że dostawcy systemów SI wysokiego ryzyka będą musieli stworzyć system zarządzania jakością i aktualizować go przez cały okres użytkowania systemu. Dodatkowo zgodnie z podejściem opartym na analizie ryzyka systemy sztucznej inteligencji wysokiego ryzyka dopuszcza się do obrotu na rynku europejskim pod warunkiem spełnienia określonych wymogów obowiązkowych i przeprowadzenia oceny zgodności *ex ante*, co oznacza, że przed wprowadzeniem systemu do obrotu lub oddaniem go do użytku istnieje obowiązek sporządzenia dokumentacji technicznej[479]. Wymogi dotyczące treści dokumentacji określono załącznikiem IV do projektowanego rozporządzenia[480]. Certyfikacja ma jednak w założeniu charakter wewnętrzny, w związku z czym podmiot wprowadzający na rynek system SI (lub jego pełnomocnik na terytorium Unii Europejskiej) będzie zobowiązany samodzielnie sporządzić deklarację zgodności oraz dokumentację techniczną[481]. Rozporządzenie stanowi jednocześnie, że organy publiczne, które oddają do użytku systemy sztucznej inteligencji wysokiego ryzyka na własny użytek, mogą przyjąć i wdrożyć zasady dotyczące systemu zarządzania jakością w ramach systemu zarządzania jakością przyjętego, stosownie do przypadku, na szczeblu krajowym lub regionalnym z uwzględnieniem specyfiki sektora oraz kompetencji i organizacji danego organu publicznego.

Zgodnie z brzmieniem wniosku po przeprowadzeniu przez dostawcę odpowiedniej oceny zgodności powinien on zarejestrować systemy sztucznej inteligencji wysokiego ryzyka, o których mowa w załączniku III do rozporządzenia, w unijnej bazie danych, która będzie

[478] *Analiza wybranych aspektów projektu…*, s. 22.
[479] Art. 11 AIA.
[480] Dokumentacja techniczna przewiduje m.in. konieczność przedstawienia: szczegółowego opisu zastosowanej procedury walidacji i testowania, w tym informacji o wykorzystanych danych walidacyjnych i danych testowych oraz ich głównych cechach; wskaźników stosowanych do pomiaru dokładności, solidności, cyberbezpieczeństwa i zgodności z innymi stosownymi wymogami określonymi w tytule III rozdziału 2, jak również potencjalnie dyskryminujących skutków; rejestrów zdarzeń generowanych podczas testów i wszystkich sprawozdań z testów opatrzonych datą i podpisanych przez osoby odpowiedzialne, w tym w odniesieniu do z góry zaplanowanych zmian, o których mowa w pkt 2 lit. f) załącznika IV, jak też szczegółowych informacji dotyczących monitorowania, funkcjonowania i kontroli systemu sztucznej inteligencji, zwłaszcza w odniesieniu do: jego możliwości i ograniczeń w zakresie skuteczności działania, w tym stopni dokładności w przypadku określonych osób lub grup osób, wobec których system ma być wykorzystywany, oraz ogólnego spodziewanego poziomu dokładności w stosunku do jego przeznaczenia; możliwych do przewidzenia niezamierzonych wyników działania i źródeł zagrożeń dla zdrowia i bezpieczeństwa, praw podstawowych i dyskryminacji w świetle przeznaczenia systemu sztucznej inteligencji; środków nadzoru ze strony człowieka wymaganych na podstawie art. 14, w tym środków technicznych wprowadzonych w celu ułatwienia użytkownikom interpretacji wyników działania systemów sztucznej inteligencji; a w stosownych przypadkach również specyfikacji dotyczących danych wejściowych.
[481] Art. 19 AIA.

142:37977

zarządzana przez Komisję w celu zwiększenia publicznej przejrzystości i kontroli oraz wzmocnienia nadzoru *ex post* sprawowanego przez właściwe organy. Natomiast w celu zachowania spójności z obowiązującymi przepisami dotyczącymi bezpieczeństwa produktów oceny zgodności systemów sztucznej inteligencji, związanymi z bezpieczeństwem elementy produktów w ramach systemu obejmującego procedury będą przeprowadzane oceny zgodności przeprowadzanej przez osobę trzecią, które już ustanowiono na podstawie odpowiednich przepisów sektorowych dotyczących bezpieczeństwa produktów. Ponowne oceny zgodności *ex ante* będą potrzebne dopiero w przypadku istotnych zmian wprowadzanych w systemach sztucznej inteligencji (zwłaszcza zmian, które wykraczają poza to, co zostało wstępnie ustalone przez dostawcę w jego dokumentacji technicznej i zweryfikowane w momencie przeprowadzania oceny zgodności *ex ante*).

Wątpliwości budzi przede wszystkim delegowanie upoważnienia do określania standardów SI, gdyż sposób redakcji treści normatywnej zakazów i nakazów może doprowadzić do ograniczonej transparentności takiego procesu i utrudnić ewentualną kontrolę sądową tych standardów. Na tym etapie prac nad projektem AIA certyfikacja ma dodatkowo wąski charakter i pomija potwierdzenie zgodności systemów z szeroko rozumianymi normami etycznymi. Nadto dostawca nie musi (z pewnymi wyjątkami) publikować dokumentacji technicznej ani jej udostępniać. Taka forma samooceny stwarza zatem ryzyko związane z brakiem odpowiednich kompetencji, a nawet woli do rzetelnej oceny systemów przez organizacje, przez co otwiera pole do nadużyć.

Kolejnym rodzajem systemów wskazanych w omawianej regulacji są systemy o ograniczonym ryzyku. Podlegają one istotnie mniejszemu zakresowi obowiązków, podczas gdy systemy o minimalnym ryzyku w ogóle nie są nimi objęte. Dostawcy systemów sztucznej inteligencji nieobarczonych wysokim ryzykiem są jednak zachęcani do wprowadzenia dobrowolnych kodeksów postępowania mających charakter dobrych praktyk. Mając jednak na uwadze dotychczasowe doświadczenia w zakresie funkcjonowania tego rodzaju dobrowolnych regulacji na gruncie RODO, należy z umiarkowanym optymizmem oczekiwać szerokiego wdrożenia kodeksów[482].

Dodatkowo w tytule IV projektu AIA wyliczono obowiązki obejmujące systemy SI szczególnego ryzyka manipulacji. Jak wynika z wniosku obejmującego projekt

[482] Na gruncie RODO w państwach członkowskich nie funkcjonują powszechnie mechanizmy certyfikacji oraz kodeksy postępowania stanowiące analogiczne narzędzie jak te proponowane w AIA.

143:39674

rozporządzenia, przepisy te będą miały zastosowanie do systemów, które: 1. wchodzą w interakcję z człowiekiem; 2. są wykorzystywane do wykrywania emocji lub określania powiązań z kategoriami (społecznymi) na podstawie danych biometrycznych; lub 3. generują treści bądź manipulują nimi (tzw. technologia *deepfake*). Zgodnie z projektem, gdy osoby fizyczne wchodzą w interakcję z systemem sztucznej inteligencji lub gdy ich emocje bądź cechy charakterystyczne są rozpoznawane za pomocą środków zautomatyzowanych, konieczne jest poinformowanie ich o tym fakcie. Dodatkowo wedle AIA, jeżeli system sztucznej inteligencji będzie wykorzystywany do generowania obrazów, dźwięków lub treści wideo, które w znacznym stopniu przypominają autentyczne treści, albo do manipulowania takimi obrazami, dźwiękami lub treściami wideo, powinien istnieć obowiązek ujawniania, że dane treści wygenerowano za pomocą środków zautomatyzowanych, z zastrzeżeniem sytuacji wyjątkowych dotyczących zgodnych z prawem celów (egzekwowanie prawa, wolność wypowiedzi). Wprowadzany mechanizm ochrony jednostki ma na celu umożliwienie dokonania świadomych wyborów, w tym umożliwienie człowiekowi wycofania się z danej sytuacji.

W tytule VIII projektu rozporządzenia określono natomiast obowiązki dostawców systemów sztucznej inteligencji w zakresie monitorowania i zgłaszania zdarzeń, a dokładniej obowiązki w zakresie monitorowania systemu po jego wprowadzeniu do obrotu oraz zgłaszania incydentów i przypadków nieprawidłowego działania związanych ze sztuczną inteligencją, a także prowadzenia dochodzeń w tych sprawach. Zgodnie z wnioskiem dostawcy SI będą zobowiązani do informowania właściwych organów krajowych o poważnych incydentach lub przypadkach nieprawidłowego działania stanowiących naruszenie obowiązków w zakresie praw podstawowych, gdy tylko się o nich dowiedzą, a także o wszelkich przypadkach wycofania systemów sztucznej inteligencji od użytkowników lub wycofania ich z rynku. Po każdym takim zdarzeniu właściwe organy krajowe przeprowadzą dochodzenie w sprawie incydentu lub nieprawidłowego działania, gromadząc wszystkie niezbędne informacje i regularnie przekazując je Komisji wraz z odpowiednimi metadanymi. Organy nadzoru rynku będą również kontrolować rynek oraz sprawdzać zgodność z obowiązkami i wymogami dotyczącymi wszystkich systemów sztucznej inteligencji wysokiego ryzyka już wprowadzonych do obrotu. Organy nadzoru rynku będą przy tym miały wszystkie uprawnienia wynikające z rozporządzenia (UE) 2019/1020 w sprawie nadzoru rynku[483]. Projekt

[483] Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2019/1020 z 20.06.2019 r. w sprawie nadzoru rynku i zgodności produktów oraz zmieniające dyrektywę 2004/42/WE oraz rozporządzenia (WE) nr 765/2008 i (UE) nr 305/2011, Dz.Urz. UE L 169, s. 1.

144:28446

rozporządzenia – odmiennie niż RODO – nie przewiduje mechanizmu kompleksowej współpracy. Istotnym problemem jest przy tym lapidarność regulacji oraz brak spójności regulacyjnej osiąganej przez pryzmat określenia uprawnień i zadań właściwych organów nadzorczych definiowanych w art. 3 pkt 43 rozporządzenia jako krajowe organy nadzorcze, organy notyfikujące i organy nadzoru rynku, które są definiowane odpowiednio w art. 4 pkt 42, 19 i 26[484]. Oznacza to, że każdy z organów nadzorczych może być właściwy do prowadzenia postępowań nadzorczych, co może doprowadzić do fragmentaryzacji nadzoru nad SI o charakterze transgranicznym[485].

Opisane obowiązki są ważne przy założeniu refleksyjnej regulacji SI, bezpośrednio nie znajdują jednak przełożenia na polepszenie sytuacji grup dyskryminowanych czy wykluczanych w automatycznych systemach decyzyjnych, aczkolwiek przy tworzeniu projektu rozporządzenia dostrzeżono również kluczowe dla interesującego mnie w tej pracy problemu *algorithmic bias* zagrożenie związane z jakością danych, na których pracują systemy SI. Rozporządzenie stanowi, że wysoka jakość danych ma zasadnicze znaczenie dla skuteczności działania wielu systemów sztucznej inteligencji, szczególnie w przypadku stosowania technik obejmujących trenowanie modeli w celu zapewnienia, aby system sztucznej inteligencji wysokiego ryzyka działał zgodnie z przeznaczeniem i bezpiecznie oraz aby nie stał się źródłem zakazanej przez prawo Unii dyskryminacji[486].

Zbiory danych do szkolenia systemów sztucznej inteligencji, ale też szerzej – wszelkich ASD, powinny być dobierane przy uwzględnieniu jakości danych, m.in. w odniesieniu do istotności, reprezentatywności, dokładności, kompletności i właściwości specyficznych dla danego obszaru zastosowań[487]. Zniekształcone dane z dużym prawdopodobieństwem doprowadzą bowiem do błędnego wnioskowania przez SI. Dodatkowo w celu opracowania systemów sztucznej inteligencji wysokiego ryzyka niektóre podmioty, takie jak dostawcy, jednostki notyfikowane i inne odpowiednie podmioty, w tym ośrodki innowacji cyfrowych, ośrodki testowo-doświadczalne i naukowcy, powinny mieć możliwość uzyskania dostępu do wysokiej jakości zbiorów danych i korzystania z nich w swoich obszarach działalności związanych z niniejszym rozporządzeniem.

[484] *Analiza wybranych aspektów projektu*…, s. 40.
[485] *Ibidem*, s. 19.
[486] Pkt 44 AIA.
[487] *Ibidem*.

145:19709

Obowiązki w zakresie testowania *ex ante*, zarządzania ryzykiem i nadzoru ze strony człowieka ułatwią również poszanowanie innych praw podstawowych poprzez zminimalizowanie ryzyka błędnych lub stronniczych decyzji podejmowanych przy wsparciu SI w obszarach krytycznych. Wobec specyfiki uczenia maszynowego postulat bezbłędności danych wydaje się jednak źle zredagowany, a przez to trudny, o ile w ogóle możliwy do realizacji w praktyce. W większości przypadków odpowiedzialne za przygotowywanie danych podmioty nie będą bowiem w stanie doprowadzić do powstania bezwzględnie prawidłowych zbiorów danych, ponieważ trudno z góry określić, jakie powinny być obiektywne warunki takiego zbioru, oraz przeprowadzić niezależną obserwację składu całej populacji[488]. Co więcej, algorytmy uczone na błędnych danych mogą powstawać celowo – np. SI, którego zadaniem będzie wykrywanie błędów, może być z rozmysłem uczony na danych niezgodnych z rzeczywistością (tj. na błędach)[489].

Z tej przyczyny w dyskusji nad projektem rozporządzenia podnosi się, że dostawcy powinni być jedynie zobowiązani do zapewnienia – z należytą starannością – poprawności danych, biorąc pod uwagę adekwatność działań względem planowanego celu wykorzystania SI[490]. Ograniczenie obowiązku do staranności, chociaż wydaje się odpowiadać na postulaty urealnienia regulacji do możliwości poznawczych człowieka, może budzić przekonanie o niedostatecznym i celowo okrojonym zakresie odpowiedzialności dostawców *big data*. Dodatkowo w celu ograniczenia błędów w stosowaniu rozporządzenia postuluje się doprecyzowanie wymogu „kompletności" danych. Jak słusznie wskazują autorzy analizy[491] pod redakcją Romana Biedy, w niektórych przypadkach wykorzystanie całości posiadanych danych (tj. „kompletnych" danych) mogłoby prowadzić do naruszenia innych przepisów AIA. Przykładowo historyczne dane dotyczące zatrudnienia mogą normalizować zachowania aktualnie nieakceptowalne (np. dyskryminację kobiet w zatrudnieniu), a wykorzystanie całości posiadanych danych do trenowania modelu mogłoby powodować zarzut niereprezentatywności czy nieadekwatności danych (art. 10 ust. 3 AIA) lub naruszenie art. 10 ust. 4 AIA[492].

Wpisane w AIA obowiązki związane z jakością danych świadczą o tym, że ustawodawca europejski dostrzega wpływ danych na stronniczość decyzji SI, jednakże aktualny kształt

[488] S. Barocas, A.D. Selbst, *op. cit.*, s. 722.
[489] *Analiza wybranych aspektów projektu…*, s. 31.
[490] *Ibidem*, s. 11.
[491] *Ibidem*, s. 32.
[492] *Ibidem*.

146:20588

regulacji pozwala na wniosek, że ochrona projektowana na gruncie AIA będzie niewystarczająca. Rozporządzenie zawiera przy tym liczne ułomności, które są przedmiotem aktualnej debaty teoretyczno-prawnej i społecznej. Dyskusja o planowanej regulacji, ale i samym jej przedmiocie jest wyjątkowo burzliwa i obszerna, przez co przedmiotem badania uczyniłam jedynie te fragmenty, które mają istotne znaczenie dla problemu dyskryminacji.

Krytycznie oceniam niedostrzeżenie nowych zagrożeń związanych z niejawnym i masowym różnicowaniem przez ASD nie tylko na podstawie cech chronionych, ale również dynamicznie dobieranych zespołów tych cech bądź innych informacji lub danych, w tym np. behawioralnych[493]. Tego rodzaju różnicowanie może zaś nie zostać uznane na naruszenie zakazu dyskryminacji na gruncie obowiązującego prawa antydyskryminacyjnego. Nadto trzeba zauważyć, że niewspółmierne rozłożenie ryzyka replikowania stronniczych decyzji, które będzie się koncentrować głównie na grupach historycznie marginalizowanych, będzie powodować błędne koło wykluczenia, a nieprecyzyjne zapisy dotyczące częściowego zakazu scoringu społecznego umożliwią omijanie regulacji. Jako niewystarczające należy także ocenić proste odwołanie AIA do treści zakazu dyskryminacji wynikającego z KPP. Realizacja ochrony na poziomie europejskim będzie bowiem wymagać rozeznania jednostki w sprawach gwarancji obywatelskich i instytucji prawnych, a ponadto posiadania przez nią zasobów finansowych i organizacyjnych umożliwiających prowadzenie nierzadko wieloletnich, kosztownych sporów sądowych.

Należy również wskazać, że projekt aktu o SI w niewielkim stopniu ogranicza zagrożenia dla praw podstawowych związane z systemami nieobjętymi załącznikiem III. Pozostawienie części systemów poza zakresem regulacji – po pierwsze, z powodu ograniczenia definicji SI, a po drugie, na skutek skoncentrowania regulacji na systemach SI wysokiego ryzyka – wobec

[493] Bettina Berendt wskazuje na obszary do dalszych rozważań, w tym konieczność uwzględnienia w regulacji: „a) przeciwdziałania uprzedzeniom i dyskryminacji (w zakresie wykrywania i zapobiegania lub przynajmniej łagodzenia wymienionych zjawisk); b) wprowadzenia obowiązku działań nakierowanych na przerwanie pętli sprzężenia zwrotnego – pętle powinny być przewidywane, wykrywane i przerywane; c) poszerzenia odpowiedzialności odszkodowawczej – prawo powinno chronić nie tylko przed zagrożeniami dla zdrowia i bezpieczeństwa, ale również obejmować szkody związane z dyskryminacją; d) zorientowania obowiązków na ochronę praw podstawowych – jeśli klasy ryzyka systemów SI mają zostać zachowane, to klasyfikacja systemu nie powinna zależeć od funkcji lub operatora, ale od oddziaływania systemu SI na prawa podstawowe; potencjalną szkodę należałoby estymować za pomocą oceny wpływu systemu na prawa podstawowe; e) należy poszerzać świadomość kategorii, jak też przeprowadzać test adekwatności, tj. w jaki sposób korzyści z takiego zastosowania przewyższają ryzyko; f) należy dopasować specyfikę metryk: muszą one odpowiadać pojęciom używanym w danej dziedzinie zastosowania; g) [konieczna jest ponadto] intensyfikacja orientacji na adresatów rozwiązań (podmioty procesów decyzyjnych)" – *eadem, The AI Act Proposal: Towards the Next Transparency Fallacy? Why AI Regulation Should Be Based on Principles Based on How Algorithmic Discrimination Works,* [w:] *Künstliche Intelligenz – Wie gelingt eine vertrauenswürdige Verwendung in Deutschland und Europa?*, red. F. Rostalski, M. Siebeck, Tübingen 2022, s. 49-51-.

147:77659

podobnego zagrożenia dyskryminacją jednostki we wszystkich rozwiązaniach algorytmicznych wydaje się trudne do uzasadnienia[494].

W mojej opinii nie wystarczy zatem zakazać w ramach AIA dyskryminacji w obecnym jej rozumieniu i odwołać się do istniejących form, aby zapobiec marginalizacji i wykluczeniu. Bardziej holistyczny proces należytej staranności powinien wykraczać poza kwestie uprzedzeń technicznych i badać konsekwencje automatyzacji podstawowych funkcji rządowych dla praw człowieka[495]. Realizując takie cele, rozporządzenie powinno zabraniać wszelkiego rodzaju punktacji behawioralnej, która nadmiernie ogranicza lub ma negatywny wpływ na prawa człowieka, w tym prawa do zabezpieczenia społecznego, odpowiedniego standardu życia, prywatności i niedyskryminacji[496].

Moje przekonania potwierdza raport z 30.11.2021 r. grupy European Digital Rights (EDRi)[497] i 119 organizacji społeczeństwa obywatelskiego, które wydały wspólne oświadczenie wzywające do przyjęcia regulacji w przedmiocie sztucznej inteligencji, stawiającej na pierwszym miejscu prawa podstawowe[498]. Niedokonanie zmian w zakresie definiowania naruszenia równości przez ASD przy jednoczesnej ograniczonej zdolności organów nadzoru do wykrywania bezprawnych praktyk w nieprzejrzystych systemach może czynić

[494] M. Veale, F. Zuiderveen Borgesius, *op. cit*.
[495] S. Chander *et al.*, *Q&A: How the EU's Flawed Artificial Intelligence Regulation Endangers the Social Safety Net*, Human Rights Watch 2021, https://www.hrw.org/news/2021/11/10/how-eus-flawed-artificial-intelligence-regulation-endangers-social-safety-net (dostęp: 4.12.2021).
[496] *Ibidem*.
[497] European Digital Rights (EDRi) to sieciowy kolektyw organizacji pozarządowych, ekspertów, adwokatów i naukowców zajmujących się obroną i rozwojem praw cyfrowych w Europie.
[498] Dokument zawiera krytyczną analizę kierunku regulacji sztucznej inteligencji oraz formułuje szeroki katalog postulatów, m.in. wskazując na konieczność: „(1) utworzenia spójnego, elastycznego i przyszłościowego podejścia do »ryzyka« systemów SI, w tym wprowadzenia procedur dotyczących aktualizacji systemów SI (w tym wprowadzenia mechanizmu aktualizacji wykazu przypadków użycia, które są uważane za niedopuszczalne); (2) ustanowienia bezwzględnych zakazów dotyczących wszystkich systemów SI stwarzających niedopuszczalne ryzyko naruszenia praw podstawowych (w tym systemów punktacji społecznej; zdalna identyfikacja biometryczna w przestrzeniach publicznie dostępnych ((przez wszystkich aktorów)); systemy rozpoznawania emocji; dyskryminacyjna kategoryzacja biometryczna; fizjonomia AI; systemy wykorzystywane do przewidywania przyszłej działalności przestępczej; systemy profilowania i oceny ryzyka w kontekście migracji); (3) wprowadzenia obowiązków użytkowników (tj. wdrażających) systemów sztucznej inteligencji wysokiego ryzyka w celu ułatwienia odpowiedzialności przed osobami, na które mają wpływ systemy sztucznej inteligencji (w tym obowiązek przeprowadzania i publikowania oceny skutków dla praw podstawowych); (4) przejrzystości publicznej (w tym obowiązek rejestrowania przez użytkowników korzystania z systemów wysokiego ryzyka w publicznej bazie danych); (5) zagwarantowania istotnych praw (takich jak prawo do niepodlegania systemom SI z naruszeniem przepisów i niedozwolonych praktyk, prawo do wyjaśnienia decyzji podjętych przy pomocy systemów SI oraz prawo do sądowego środka odwoławczego) i zadośćuczynienia dla osób dotkniętych przez systemy sztucznej inteligencji; (6) spełnienia wymogu dostępności w całym cyklu życia SI; (7) zrównoważonego rozwoju i ochrony środowiska przy opracowywaniu i użytkowaniu systemów SI; (8) ulepszonych i przyszłościowych standardów dla systemów SI" – *An EU Artificial Intelligence Act for Fundamental Rights. A Civil Society Statement*, 2021, https://edri.org/wp-content/uploads/2021/12/Political-statement-on-AI-Act.pdf (dostęp: 4.12.2021).

148:10473

projektowany akt nieistotnym dla poprawy sytuacji grup marginalizowanych historycznie. Należy przy tym oczekiwać dalszych prac nad projektem, co może się przyczynić do zmiany jego brzmienia.

Praktycznym utrudnieniem dla realizacji ochrony przed dyskryminacją w toku stosowania SI jest jednak niewskazanie w AIA drogi do zakwestionowania skutków prawnych takiej dyskryminującej decyzji systemu SI. Z perspektywy obywatela, o ile decyzja ma charakter dyskryminujący, powinna zostać przewidziana droga do jej zakwestionowania. W tym zaś przypadku zakres przedmiotowy AIA nie koncentruje się na bezprawności decyzji jako takiej, a na funkcji i sposobie wykorzystania danego systemu, z którego decyzja pochodzi. Sama penalizacja operatorów systemów SI będzie oczywiście niewystarczająca dla zmiany sytuacji prawnej pokrzywdzonej jednostki lub społeczności. Dlatego też kolejnym krokiem i elementem całego programu związanego z budowaniem godnej zaufania sztucznej inteligencji jest proponowana przez Komisję Europejską harmonizacja krajowych przepisów dotyczących odpowiedzialności za SI. Celem regulacji jest ujednolicenie zasad w celu poprawy sytuacji jednostki poszkodowanej działaniem SI. Dyrektywa w sprawie odpowiedzialności za sztuczną inteligencję[499] ma ułatwić dochodzenie roszczeń odszkodowawczych z tytułu odpowiedzialności cywilnej dzięki określeniu m.in. zasady dostępu do informacji i ciężaru dowodu. Jest to ważny krok w kierunku urzeczywistnienia potrzebnej ochrony. Regulacja przewiduje bowiem uprawnienie dla podmiotów prawnych i fizycznych do występowania z roszczeniami cywilnymi w związku z funkcjonowaniem systemów sztucznej inteligencji. Na jej podstawie przyznane zostanie roszczenie informacyjne m.in. w zakresie funkcjonowania systemu SI w celu zidentyfikowania przyczyn szkód.

Projekt spotkał się jednak z bardzo rozbieżnymi opiniami. Podczas gdy rynek obawia się ograniczenia innowacyjności, organizacje społeczne wskazują, że zakres regulacji jest zbyt wąski[500]. Zanim wniosek stanie się prawem, a na jego podstawie państwa członkowskie będą musiały podjąć krajowe działania legislacyjne, musi jeszcze pokonać kilkuletnią drogę unijnego procesu prawodawczego. W tym czasie – podobnie jak w wypadku AIA – zostanie on z pewnością wielokrotnie zmieniony i poddany intensywnemu lobbingowi ze strony firm

[499] Wniosek: dyrektywa Parlamentu Europejskiego i Rady w sprawie dostosowania przepisów dotyczących pozaumownej odpowiedzialności cywilnej do sztucznej inteligencji (dyrektywa w sprawie odpowiedzialności za sztuczną inteligencję), COM(2022) 496 final.

[500] Projektowana regulacja nakłada na konsumentów odpowiedzialność za udowodnienie, że system sztucznej inteligencji wyrządził im krzywdę lub że twórca sztucznej inteligencji dopuścił się zaniedbania.

149:66576

technologicznych, przez co trudno na tym etapie ostatecznie przesądzić o jego znaczeniu dla ochrony przed dyskryminacją.

3. Logika regulacji SI

Unia Europejska dąży do tego, aby AIA – podobnie jak RODO – okazał się nowym światowym złotym standardem regulacji sztucznej inteligencji. I choć grupy eksperckie i polityczne dostrzegły konieczność minimalizacji ryzyka dyskryminacji, projekt AIA nie egzemplifikuje odmienności w różnicowaniu algorytmicznym, a istotny ciężar przenosi na określenie obszarów, w których dyskryminacja może się ujawnić (np. zatrudnienie, opieka zdrowotna nadzór państwowy, identyfikacja biometryczna). Regulacja ma mieć przy tym charakter refleksyjny. Prawodawca identyfikuje poziomy ryzyka od akceptowalnego do nieakceptowalnego. Na poziomie rozporządzenia zmienia się zatem logika ustawodawcy, który określa istotne dla zakazu dyskryminacji sfery życia społecznego narażone w szczególny sposób na bezprawne różnicowanie. AIA nie może jednak – z przyczyn ustrojowych – modyfikować czy uzupełniać rozwiązań przyjętych w Karcie Praw Podstawowych w zakresie rozumienia samego zakazu lub form dyskryminacji nim objętych. Z pewnością zaś obecne brzmienie rozporządzenia nie będzie dostatecznie ograniczało wpływu ASD na opisaną w rozdziale III niniejszej pracy marginalizację i wykluczenie grup historycznie uciśnionych. Marginalizacja ta nie będzie miała bowiem miejsca jedynie w konsekwencji użycia wobec przedstawicieli grup historycznie wykluczonych sztucznej inteligencji w obszarach uznanych za niedopuszczalne. Ewentualne spory na tle zakresu pojęcia dyskryminacji w kontekście nieprzejrzystych procesów różnicowania nadal będą podlegały dyskrecjonalnej decyzji sędziów (w tym na poziomie krajowym) i będą musiały się zmagać z niewykształceniem adekwatnego zakresu pojęcia prawnego zakazu dyskryminacji wobec specyfiki różnicowania algorytmicznego.

Dodatkowo projektowane rozporządzenie pomija istotne praktyczne utrudnienia w stosowaniu prawa antydyskryminacyjnego, nie odpowiadając na podstawowe pytania, np. według jakiego prawa materialnego winien być rozstrzygany konflikt, który powstaje między ofiarami dyskryminacji a dostawcami SI, i który sąd będzie miał w tym zakresie jurysdykcję, a także jaka będzie w takim procesie rola państwowych organów nadzorczych[501]. Kwestie te być może

[501] R. Xenidis, L. Senden, *op. cit.*

150:10437

zostaną wyjaśnione w kolejnych aktach prawnych, jednak – jak widać na przykładzie projektu dyrektywy w sprawie odpowiedzialności za sztuczną inteligencję – proces legislacyjny może być długotrwały oraz narażony na rozliczne naciski i praktyki lobbingowe branży technologicznej.

Narzędzia regulacyjne w domenie publicznej tradycyjnie opierały się na świadomym projektowaniu norm, podczas gdy obecnie coraz częściej odchodzi się od tego modelu w kierunku prawa aktywnie reagującego na zmieniające się otoczenie, mającego potencjał podążania i uczenia się na podstawie relacji społeczno-politycznych[502]. Coraz śmielej do opracowywania tych norm stosuje się również metodologie *data science*[503]. Z funkcjonalnego punktu widzenia zarówno prawo tradycyjne, jak i technologia mogą równocześnie, ale też niezależnie działać jako mechanizmy regulacyjne, które dążą do podporządkowania ludzkiego postępowania pewnym regułom[504], przy czym technologia na obecnym etapie rozwoju zasadniczo nie jest zobligowana do reprezentowania konkretnych standardów moralnych czy wartości etycznych. Dodatkowo prawo jest tworzone w sferze publicznej zgodnie z ustrojowymi zasadami, regulacje wpisane w technologie są niejawne i niepewne pod kątem treści oraz stałości reprezentowanych norm, a sama droga do ich zakodowania jest nieprzejrzysta dla laików. Wykorzystanie automatycznych systemów decyzyjnych, zwłaszcza gdy system uniemożliwia uzasadnienie decyzji w ramach wykonywania władzy publicznej, stawia tymczasem pod znakiem zapytania podstawowe dogmaty państwa prawa.

Statystyczne potwierdzenie prawdziwości decyzji wynikające z analizy danych może jednocześnie naruszać idee jakościowe, w tym równość, uczciwość czy skuteczność. Mechanizmy wprowadzone w projekcie rozporządzenia nie przystają bowiem do wyzwania zagwarantowania praw podstawowych jednostce w konfrontacji z problemem algorytmicznego różnicowania. Akt w sprawie sztucznej inteligencji reguluje zarówno narzędzia nadzoru i kontroli, jak i procedury certyfikacji czy poziomy ryzyka systemów SI, jednak nie podejmuje próby zrozumienia mechanizmu różnicowania przez algorytmy. Akt znajdzie przy tym zastosowanie tylko w zakresie przedmiotu regulacji, której granice będą wyznaczały definicja legalna SI, typologia i klasyfikacja ryzyka, a także użycie systemu SI w konkretnej sferze życia człowieka.

[502] M. Pichlak, K. Gaczoł, *op. cit.*
[503] E. Bayamlıoğlu, R. Leenes, *op. cit.*
[504] *Ibidem*.

151:10663

Projektowana regulacja określa przy tym adresatów nowych obowiązków, ograniczając w ten sposób sam przedmiot regulacji. Ma ona bowiem unormować zachowanie ludzkie, pozostawiając bez wystarczającej odpowiedzi niebezpieczeństwa algorytmicznego procesu decyzyjnego związanego z jego nieprzejrzystością technologiczną, a zatem samym rdzeniem mechanizmu uczenia maszynowego. Kodyfikowanie algorytmów przy wykorzystaniu języka naturalnego za pomocą tradycyjnych narzędzi regulacyjnych, nawet o potencjale adaptacyjnym – jak AIA czy RODO – będzie miało skutek co najwyżej na poziomie obowiązków człowieka operatora (tj. od etapu tworzenia do wykorzystania SI). Ludzie jednak częstokroć nie mają wiedzy o działaniu SI ani realnego wpływu na nie, nie tylko ze względu na uczenie algorytmu, ale też rozdzielenie procesu na tworzenie i uczenie z wykorzystaniem danych algorytmu. Skoro więc systemy SI cechuje możliwość podejmowania decyzji w sposób niekontrolowany, obrana droga regulacji zakazu dyskryminacji przez ASD może być nieudolna narzędziowo. Człowiek przy stosowaniu dostępnych mu pozacyfrowych narzędzi, działając w granicach własnego poznania, nie jest bowiem w stanie realizować efektywnie kontroli nad systemami SI z powodu ograniczeń kognitywnych[505]. Istniejący porządek prawny został wszak skonstruowany dla relacji symetrycznych, w których podmioty prawne są sobie równe w zdolnościach poznawczych, a ewentualne nadużycia są penalizowane[506]. Natomiast asymetria przejawia się tutaj w nieskuteczności języka naturalnego jako narzędzia regulacji i przekazaniu sporów pod ograniczone poznawczo sądy ludzkie. Skoro tradycyjne ujęcie i rozumienie funkcjonowania prawa, a przede wszystkim tradycyjna technika legislacyjna są niewystarczające do uregulowania SI w sposób dający się wyegzekwować, to należy rozważyć, czy forma i instrumenty regulacji nie powinny ściślej podążać za jej funkcją i przedmiotem[507].

Przeciwdziałanie dyskryminacji w automatycznych procesach decyzyjnych wymaga konsultacji i współtworzenia przyszłych regulacji ze specjalistami z zakresu IT. Istotną barierą dla prawników i organów nadzoru jest bowiem nieprzejrzystość epistemologiczna przejawiającą się w braku dostępu do wewnętrznej logiki procesu różnicowania SI. Rozdział kompetencji powinien przy tym umożliwić prawnikom interpretowanie reguł społecznych wpisanych w reguły prawne, a programistom – określenie architektury technologii przez

[505] Zygmunt Bauman pisze o uniemożliwiającym „nadmiarze informacji"; zob. *idem*, *Życie na przemiał*, Kraków 2005.
[506] Art. 84, art. 86 i art. 388 KC.
[507] D. Szostek, *Sztuczna inteligencja a kody. Czy rozwiązaniem dla uregulowania sztucznej inteligencji jest* Smart Contract *i* Blockchain*?*, [w:] *Prawo sztucznej inteligencji*…, Legalis/el.

152:11084

konstrukcje kodu umożliwiającego zarówno funkcjonowanie w cyberprzestrzeni zgodnie z regułami prawa, jak i nawet egzekwowanie prawa[508].

Ograniczenie regulacji jedynie do postulatów równościowych, których adresatami są dostawcy systemów SI, przy jednoczesnym niezrozumieniu mechanizmu i skali zagrożenia nie pozwala dokonywać racjonalnych wyborów na poziomie narzędzi regulacyjnych. AIA pozwala wprawdzie na adaptację przepisów do zmieniających się stosunków społeczno-gospodarczych, jednak koncentruje się na tradycyjnych instrumentach prawa i klasycznym rozumieniu procesów decyzyjnych. Przyczynami zaistnienia omawianej bariery są: nowość dziedziny, tempo jej doskonalenia i odmienności w zakresie sposobu funkcjonowania mechanizmów decyzyjnych algorytmów polegających na uczeniu maszynowym i przetwarzaniu dużych zbiorów danych w porównaniu do podejmowania decyzji przez ludzi. SI charakteryzuje się bowiem heurystycznym i dynamicznym podejściem do interpretowania zebranych ustrukturyzowanych lub nieustrukturyzowanych danych oraz wyciąganiem wniosków z wiedzy o charakterze statystycznym i zaobserwowanym w danych[509]. Tego typu technologia, optymalizując procesy, dotyka fundamentalnych wartości w stosowaniu sprawiedliwego systemu podejmowania decyzji. W takim przypadku należy rozważyć, czy formuła umożliwiająca samodzielne uczenie się nałożonych zasad, nakazów oraz zakazów, a więc podstawowych dla niej paradygmatów implementowanych w kodach algorytmicznych, uprzednio przygotowanych przez ludzi, nie byłaby bardziej efektywna do zamierzonego celu[510]. Na potrzeby dalszej analizy ten typ prawa nazywam technoregulacjami.

Droga stworzenia i implementowania do istniejącego porządku prawnego technoregulacji z pewnością nie jest pozbawiona wątpliwości natury ustrojowej i etycznej, jednak kierunek rozwoju technologii wymusza zmianę podejścia do efektywnych narzędzi prawnych. Niestety jeśli uznamy ten kierunek za zasadny, pojawi się problem tworzenia prawa niewynikającego jedynie z konstytutywnej inicjatywy ustawodawczej, a z normatywnego oddziaływania na treść regulacji automatycznego procesu decyzyjnego. Ramy takich technoregulacji będą kreowane przez programistów tworzących algorytm, a sam algorytm na skutek procesu uczenia (tzn. wniosków płynących z analizy danych treningowych) będzie te ramy wypełniał, a być może

[508] D. Szostek, *LegalTech. Czyli jak bezpiecznie korzystać z narzędzi informatycznych w organizacji, w tym w kancelarii oraz dziale prawnym*, Warszawa 2021.
[509] D. Szostek, *Is the Traditional Method of Regulation (the Legislative Act) Sufficient to Regulate Artificial Intelligence, or Should It Also Be Regulated by an Algorithmic Code?*, „Bialystok Legal Studies" 26, 2021, nr 3, s. 43–60.
[510] *Ibidem*.

153:59727

nawet adaptował. Z uwagi na systemy czarnoskrzynkowe i nieprzejrzystość analizy pomiędzy poszczególnymi warstwami adresat decyzji nie będzie mógł rozróżnić, która część normatywna decyzji jest zamierzona (można ją wywnioskować z ram), a która jest jedynie efektem ubocznym w postaci nieprzewidzianych lub wtórnych skutków procesu automatycznego[511]. Takie technoregulacje z racji stopnia złożoności i nieprzejrzystości będą mogły podlegać jedynie algorytmicznemu nadzorowi, co w dalszym zakresie będzie powodować odpływ tradycyjnie rozumianego władztwa publicznego w kierunku technologii i jej dostawców. Ludzkie postrzeganie kontroli algorytmów nad prawidłowym stosowaniem technoregulacji może być przez to równie, a może nawet bardziej nieprzejrzyste niż obecny automatyczny proces decyzyjny. To zaś otwiera dyskusję o aktualności podstawowych imperatywów dotyczących wykonywania władzy publicznej, mianowicie prawa obywateli do przejrzystości, rozumianej jako dostępność i zrozumiałość informacji, a także o konieczności uzasadnienia podejmowanych decyzji dotyczących praw i obowiązków jednostki.

W systemach całkowicie zautomatyzowanych, opierających się na algorytmach oraz na systemach uczących się i sztucznej inteligencji prawidłowość decyzji automatycznej nie zostanie efektywnie skontrolowana przez człowieka[512]. Bez względu na dalszy kierunek rozwoju narzędzi prawnych skala interakcji społecznych zarządzanych obecnie przez automatyczne systemy decyzyjne sprawia, że w coraz większym stopniu technologia powinna bezpośrednio egzekwować reguły leżące u podstaw państwa demokratycznego[513]. Reguły te są jednak trudne do zaimplementowania do omawianych systemów, albowiem język programowania jest sformalizowany, a przez to nieprzystający do prawa pisanego językiem naturalnym[514].

Ustanowienie reguł w sposób możliwy do odczytania przez algorytmy przyczyniłoby się tymczasem do polepszenia jakości decyzji. Taką drogę dostrzegał Lawrence Lessig, który traktował prawo jak kod[515]. Badacz chciał, aby system prawa składał się z „puzzli", które

[511] E. Bayamlıoğlu, R. Leenes, *op. cit*.
[512] M. Kibil, *op. cit*.
[513] E. Mokhtarian, *The Bot Legal Code: Developing a Legally Compliant Artificial Intelligence*, „Vanderbilt Journal of Entertainment & Technology Law" 21, 2018, s. 145–207.
[514] S. Hassan, P. De Filippi, *The Expansion of Algorithmic Governance: From Code Is Law to Law Is Code*, „Field Actions Science Reports" 17, 2017, s. 88–90.
[515] Lessig stwierdza, że „regulatorem jest kod – oprogramowanie i sprzęt, które sprawiają, że cyberprzestrzeń jest taka, jaka jest. Ten kod, czy też architektura, określa warunki, na jakich doświadcza się życia w cyberprzestrzeni. Określa on, jak łatwo jest chronić prywatność lub cenzurować wypowiedzi. Ma wpływ na to, kto co widzi lub co jest monitorowane, a co niewidoczne. Kod reguluje cyberprzestrzeń w sposób, którego nie sposób dostrzec, dopóki nie zaczniemy rozumieć natury tego kodu. Decyduje o tym, czy dostęp do informacji jest powszechny. W

154:96917

można by łączyć z sobą i kształtować m.in. w cyberprzestrzeni. Automatyzacja podejmowania decyzji bezsprzecznie stwarza jednak problemy spójności etycznej między ludźmi a algorytmami[516]. Zasady moralne obowiązujące w społeczeństwach są bowiem niejednoznaczne i niejednolite, a w związku z tym trudne do przełożenia na uniwersalnie stosowany kod algorytmu[517]. Nadto nie istnieje żaden powszechnie przyjęty system etyki, który uniwersalnie można by do tego celu uznać za obowiązujący. Formalizacja etyki w predykcji z wykorzystaniem algorytmów uczenia maszynowego musiałaby przy tym odpowiadać na problem dążenia do osiągnięcia równości materialnej przy uwzględnieniu kwestii społecznych, takich jak nierówny dostęp do zasobów i uwarunkowania społeczne[518].

Pomimo wielu formalnych, matematycznych definicji sprawiedliwości zaproponowanych w ciągu ostatnich lat problem formalizacji sprawiedliwości jest wciąż otwarty[519]. Pojęcie sprawiedliwości w języku naturalnym różni się bowiem w zależności od obszaru geograficznego, kraju, kultury i dziedziny zastosowania, przez co sprawiedliwości nie da się sprowadzić do prostej, samodzielnej definicji, a z pewnością występują również duże trudności w jej matematycznym skodyfikowaniu[520]. Rozumienie tego pojęcia ma dynamiczny wymiar społeczny, przez co nie będzie łatwo odzwierciedlić je w ramach analizy statystycznej.

Podobne trudności pojawiają się w przypadku norm prawnych, które z natury są abstrakcyjne[521]. Podjęcie takiej próby wymagałoby zmiany sposobu myślenia o normie prawa jako bycie treściowym zakodowanym w formie gramatycznej, jaką jest przepis prawny[522]. To

miarę jak ten kod się zmienia, zmienia się też natura cyberprzestrzeni" – *idem*, *Code: And Other Laws of Cyberspace, Version 2.0*, New York 2006, s. 121.

[516] M. Turilli, *Ethical Protocols Design*, „Ethics and Information Technology" 9, 2007, nr 1, s. 49–62.

[517] E. Mokhtarian, *op. cit.*, s. 173.

[518] P. Gajane, M. Pechenizkiy, *On Formalizing Fairness in Prediction with Machine Learning*, Cornell University, New York 2018, https://arxiv.org/pdf/1710.03184.pdf (dostęp: 4.12.2021).

[519] Sam Corbett-Davies i Sharad Goel twierdzą, że „[w] rodzącej się dziedzinie sprawiedliwego uczenia maszynowego trzy formalne definicje sprawiedliwości zyskały na znaczeniu: (1) antyklasyfikacja, co oznacza, że chronione atrybuty – takie jak rasa, płeć – i ich nosiciele nie są wprost wykorzystywani do podejmowania decyzji; (2) parytet klasyfikacji, co oznacza, że wspólne miary skuteczności predykcyjnej (np. współczynniki wyników fałszywie dodatnich i fałszywie ujemnych) są równe we wszystkich grupach zdefiniowanych przez chronione atrybuty; oraz (3) kalibracja, co oznacza, że w zależności od oszacowań ryzyka wyniki są niezależne od chronionych atrybutów. Wszystkie trzy z tych definicji sprawiedliwości mają istotne ograniczenia statystyczne" – *eidem*, *The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning*, Cornell University, New York 2018, s. 1, https://doi.org/10.48550/arXiv.1808.00023 (dostęp: 4.12.2021).

[520] M. Schäfer, D.B.M. Haun, M. Tomasello, *Fair Is Not Fair Everywhere*, „Psychological Science" 26, 2015, nr 8, s. 1252–1260.

[521] Norma prawna to zrekonstruowana z przepisów prawnych abstrakcyjno-generalna norma postępowania ustanowiona lub uznana przez prawodawcę, formułująca wzorzec wymaganego (nakazanego, zakazanego) bądź dozwolonego zachowania oraz ewentualnie podająca ujemne konsekwencje jego niedochowania; tak: G. Maroń, *Wstęp do prawoznawstwa*, Rzeszów 2011, s. 89–90.

[522] B. Liżewski, *System prawa a porządek prawny. Między konstrukcją normatywną a prawem w działaniu*, Warszawa 2021.

155:17724

zaś miałoby fundamentalny wpływ na możliwość rekonstrukcji normy. Nadto należy nadmienić, że treść przepisów zawiera odwołania do klauzul generalnych i zwrotów szacunkowych, a także do procesu interpretacji prawa, w którym istotną rolę odgrywają sądownictwo i doktryna[523].

Za Bartoszem Liżewskim można zatem przyjąć, że nawet zdekodowanie normy prawnej do algorytmu nie wykluczy problemów z uniwersalnym stosowaniem regulacji. Przede wszystkim utrudnieniem będzie to, że prawo może podlegać różnej wykładni w zależności od istniejącego układu społecznego, ekonomicznego czy politycznego. Samo stosowanie prawa wpisanego w kod musiałoby więc ewoluować. Możliwym kierunkiem, który postuluje się w literaturze, jest skonstruowanie algorytmów SI na podstawie danych zapisanych w dostępnych repozytoriach opartych np. na blockchainie[524]. Technologia *blockchain*[525] pozwoliłaby zapewnić wiarygodność zapisu i jego niezmienność, a w praktyce byłaby dowodem, że dane przekazywane do ASD są poprawne, co łącznie umożliwiłoby kontrolę procesu decyzyjnego przez człowieka[526]. Wybierając tę drogę, ustawodawca będzie się jednak musiał zmierzyć również z problematyką podmiotowości prawnej SI. Zagadnienie to z powodu zakresu i obszerności problemu nie będzie jednak przedmiotem dalszej analizy w niniejszej rozprawie.

Na obecnym etapie trudno wyobrazić sobie rychłe wpisanie w kod algorytmu norm antydyskryminacyjnych wobec istniejącej złożoności i wieloaspektowości treści zakazu bezprawnego różnicowania. W statystyce i uczeniu maszynowym wciąż pojawiają się nowe techniki, które są rozwijane w odpowiedzi na obserwowane problemy związane z dyskryminacją automatyczną. Dodatkowo nowe teorie matematyczne, takie jak „sprawiedliwość poprzez świadomość”[527], wspierają gwarancje statystycznego parytetu. Z pewnością techniki te będą udoskonalane i z czasem powszechnie wdrażane w algorytmy. Koncepcja regulatora SI jako technoregulacji na tym etapie wydaje się jednak futurystyczna, jako że prognozy co do ostatecznych kierunków rozwoju technik regulacyjnych automatycznych systemów decyzyjnych są niepewne.

[523] *Ibidem*.
[524] K. Werbach, *The Blockchain and the New Architecture of Trust*, Cambridge, MA 2018.
[525] *Blockchain* to tzw. łańcuch bloków, będący technologią uważaną za odporną na manipulacje. Jej zadaniem jest przechowywanie i przesyłanie informacji. Kontrola odbywa się wyłącznie przez zestaw niezmiennych i niezniszczalnych reguł osadzonych w kodzie źródłowym. Warto dodać, że *blockchain* leży u podstaw m.in. bitcoina.
[526] D. Szostek, *Is the Traditional Method…*
[527] J. Polonetsky, O. Tene, *Taming the Golem: Challenges of Ethical Algorithmic Decision Making*, „North Carolina Journal of Law & Technology” 19, 2018, nr 1, s. 125–173.

156:94084

Jak zauważa Kate Crawford w książce *Atlas of AI*[528], przez próbę rozwiązania problemów narzędziami technologicznymi możemy utracić z pola widzenia istotę problemu. Badaczka widzi rzeczywistą przyczynę takiego zjawiska w masowym wykorzystywaniu mechanizmu klasyfikowania, czyli *de facto* uprzedmiotawiania ludzi na wzór scjentystycznej taksonomii[529]. Jeśli zaś wykluczenie algorytmiczne będziemy traktować wyłącznie jako błąd algorytmu, który da się naprawić środkami technicznymi, to logika leżąca u podstaw źródła zjawiska marginalizacji społecznej pozostanie poza zasięgiem regulacji i poza odpowiedzią prawa. Automatyzacja procesów decyzyjnych, jak zauważa Crawford, ma przede wszystkim wymiar ekonomiczny, a automatyzacja i analiza wielkich zbiorów danych na skalę masową są uwikłane w relacje władzy istniejące w społeczeństwie[530]. Stworzenie systemu regulacyjnego opartego na algorytmach jest przy tym jedną z ciekawszych i bardzo inspirujących dróg ograniczenia problemu dyskryminacji automatycznej, jednak na początkowym etapie takie rozwiązania będą zależały od relacji władzy na styku polityczno-gospodarczym. Następnie można oczekiwać, że wdrożenie takich rozwiązań na poziomie międzynarodowym będzie koszto- i czasochłonne. Wobec społecznej wagi problemu automatycznej dyskryminacji bierne oczekiwanie na postęp technoregulacji nie jest więc zasadne.

W tym rozdziale przeanalizowałam założenia i drogę, jaką do tej pory przeszedł regulator w obszarze sztucznej inteligencji z perspektywy realizacji zakazu dyskryminacji. W toku analiz i prac legislacyjnych na poziomie Unii Europejskiej dostrzeżono zagrożenia systemów SI prowadzące do wzmocnienia nierówności społecznych, na skutek czego prawa podstawowe oraz procesy i wartości demokratyczne stały się narażone na istotne ryzyko[531]. Projektowany AIA ma w związku z tym w założeniu mieć walor adaptacyjny wobec stałego ryzyka naruszeń, jednak narzędzia w nim przewidziane mogą się okazać niedostosowane do sedna problemu ASD.

Ukazana tu analiza prowadzi mnie do wniosku, że projektowane regulacje nie zawierają prawniczej refleksji nad adekwatnością obecnych form dyskryminacji wobec specyfiki

[528] K. Crawford, *Atlas of AI*, New Haven-London 2021.
[529] O epistemologicznej przemocy taksonomii pisał już w latach 60. i 70. ubiegłego wieku Michel Foucault (do którego Crawford się odwołuje), choć źródła jego myśli sięgają Nietzschego i innych filozofów z tzw. szkoły podejrzeń.
[530] Crawford wskazuje również, że ośrodki naukowe i biznesowe tworzące sztuczną inteligencję nie tylko same w sobie cieszą się dużą władzą, ale pozostają także w coraz bliższych relacjach z władzą państwową. Taka symbioza prowadzi zaś do sytuacji, w której scentralizowane mechanizmy kontroli stopniowo wypierają ze sfery publicznej inne formy współżycia zbiorowego. Wiąże się to zarówno z ekonomicznym znaczeniem ograniczeń regulacyjnych dla rozwoju biznesu cyfrowego, jak i z ogromnym lobby gigantów technologicznych.
[531] F. Zuiderveen Borgesius, *Price Discrimination, Algorithmic Decision-Making, and European Non-Discrimination Law*, „European Business Law Review" 2020, s. 4.

157:21448

różnicowania algorytmicznego. Samo odwołanie w projekcie AIA do konieczności realizacji ochrony antydyskryminacyjnej istniejącej na gruncie aktów prawnych omówionych w rozdziale III nie jest wystarczające. Przez taki zabieg, tzn. wprowadzenie zakazu dyskryminacji przy jednoczesnym niedopasowaniu treści zakazu do specyfiki różnicowania automatycznego, istnieje ryzyko uchwalenia pozornej, niewystarczającej i nieskutecznej ochrony. Jak wskazałam w rozdziale III, grupy niekorelujące z cechami chronionymi nie korzystają obecnie z ochrony prawnej, chyba że w konkretnym przypadku dojdzie do uzupełnienia katalogu cech poprzez odpowiednie orzecznictwo lub zastosowanie w toku procesu sądowego takich regulacji jak RODO czy prawo do prywatności. Taki kształt i forma realizowania postulatów równościowych są dysfunkcyjne i niepewne z perspektywy konieczności zapewniania ochrony ogółowi społeczeństwa.

Niezależnie od tego należy podkreślić, że projektowany akt ze względu na zakres przedmiotowy regulacji obejmuje tylko część decyzji automatycznych – takich, które zostały zakwalifikowane jako efekt stosowania kwalifikowanych systemów automatycznych, tj. algorytmów sztucznej inteligencji. Z powodu tych negatywnych dla przedmiotu badania wniosków w trzecim podrozdziale tego rozdziału podjęłam rozważania nad możliwością stworzenia i zakodowania w algorytmach prawa w postaci tzw. technoregulacji, umożliwiających ograniczenie zjawiska dyskryminacji. Wieloaspektowość problemów wynikających z powiązanej z omawianą kwestią rewolucji w zakresie obowiązujących dogmatów ustrojowych na gruncie procesu tworzenia i kontroli nad technoregulacjami oraz z działania silnego lobby technologicznego prowadzi mnie do przekonania, że tematyka technoregulacji – chociaż z pewnością będzie w przyszłości odgrywać coraz większą rolę w regulowaniu algorytmów – obecnie nie może być wystarczającym narzędziem do skutecznego przeciwdziałania zjawisku dyskryminacji automatycznej.

158:75177

V. Automatyczna dyskryminacja

Frank Easterbrook[532] w latach 90. XX w. powiedział, że „najlepszym sposobem na poznanie prawa właściwego dla wyspecjalizowanych przedsięwzięć jest studiowanie zasad ogólnych". Wobec istniejącego problemu, którego rozwiązanie musi pokonać istotne przeszkody na gruncie obowiązujących i projektowanych regulacji prawnych, konieczne jest przeprowadzenie analizy z punktu widzenia relacji między treścią zakazu dyskryminacji a rzeczywistością stosunków dominacji. Z tego względu w ostatnim rozdziale niniejszej pracy rozważę ewentualne kierunki redefiniowania pojęcia i form dyskryminacji oraz możliwość realizacji postulatów równościowych w zmieniającym się otoczeniu technologicznym.

W okresie rozwoju państwa zcentralizowanego gwarancje wolności i równości miały zapewnić sądowa kontrola konstytucyjności stanowionego prawa i decyzji administracji oraz powszechny dostęp do sądów. W ostatnich latach ten paradygmat uległ jednak zmianie. Jak wiadomo, teoria konstytucyjna ujmuje władzę jako historycznie przypisaną organom publicznym, które domyślnie mają monopol na środki przemocy w ramach umowy społecznej[533]. Obecnie władza publiczna rozmywa się, ustępując miejsca realnemu zwiększeniu się jej po stronie gospodarczych gigantów technologicznych. Dzieje się to jednak nie ze względu na koncyliacyjne wypracowanie nowego modelu społeczeństwa obywatelskiego, a z racji stałego poszukiwania nowych metod racjonalizacji rynków zbytu. Jednocześnie algorytmy automatycznych systemów decyzyjnych, z racji swojej efektywności, dominują w sektorze prywatnym, choć ich rola w domenie publicznej stale się zwiększa. Procesy klasyfikowania użytkowników, wyborców, klientów, pacjentów czy konsumentów, których skutkiem jest przydzielenie szans lub ich odebranie, wtapiają się w nową rzeczywistość społeczną[534]. ASD mają przy tym zdolność do niespotykanego dotąd na skalę masową grupowania ludzi w nieoczekiwany sposób. Efektem automatycznego procesu decyzyjnego niestety może być zwiększenie skali procesu wykluczenia grup niedoreprezentowanych, a także dyskryminacja lub marginalizacja grup, które niekoniecznie identyfikują się z grupami prawnie chronionymi czy też odpowiadają kryteriom zaliczania się do takich społeczności. Równocześnie mechanizmy prowadzące do dyskryminacji i wykluczenia stają się dla człowieka nieprzejrzyste, niejawne i niedostrzegalne w procesie decyzyjnym. Wraz z

[532] Sędzia Sądu Apelacyjnego Stanów Zjednoczonych dla Siódmego Okręgu.
[533] O. Pollicino, G. De Gregorio, *op. cit.*, s. 15.
[534] G.C. Bowker, S.L. Star, *Sorting Things Out. Classification and Its Consequences*, Cambridge, MA 2000.

159:43164

automatyzacją zwiększa się przy tym potencjał statystycznego uzasadniania słuszności i prawdziwości takich decyzji, co może doprowadzić do racjonalizacji dyskryminacji[535].

Eksploracja zbiorów danych w automatycznych systemach decyzyjnych zawsze jest przy tym formą statystycznej (a więc pozornie racjonalnej) dyskryminacji[536]. Uzasadnieniem stojącym za taką formą różnicowania jest możliwość wskazania obiektywnej podstawy, na której oparta jest kategoryzacja ludzi, a nadto ekonomiczna opłacalność ograniczenia nietrafnych decyzji. Pomimo krytyki podejścia opartego na bezrefleksyjnym dopuszczeniu stosowania ASD we wszystkich obszarach życia społeczeństwo i władza mogą nie dostrzec wiążącej się z nimi niesprawiedliwości, przez co we właściwym czasie może zabraknąć odpowiedniego sprzeciwu. Należy mieć przy tym na względzie, że w odróżnieniu od tradycyjnych procesów decyzyjnych negatywne skutki eksploracji danych przeważnie nie wynikają z ludzkich uprzedzeń, świadomych lub nieświadomych, a ponadto nie są widoczne w sposób znany z historycznej walki o równość[537]. To zaś czyni dyskryminację problemem aktualnej treści pojęcia nastawionego na ograniczenie rzeczywistej i szkodliwej społecznie dominacji.

1. Redefiniowanie zakresu i form dyskryminacji

W procesach decyzyjnych ASD następuje kategoryzacja ludzi na podstawie danych i informacji mogących wymykać się definicji cech chronionych prawem antydyskryminacyjnym, a przez to mogących pozostawać poza zakresem zakazu dyskryminacji. Uzupełnieniem odpowiedniej regulacji pod pewnymi względami mogą być co prawda przepisy szczegółowe (tj. RODO, AIA czy prawo do prywatności), jednak opisywane mechanizmy ochronne nie dotykają sedna problemu, są fragmentaryczne i trudne do zastosowania w kontekście specyfiki i nieprzejrzystości analizy danych w systemach automatycznych. Procesy decyzyjne, w których algorytmy pośrednio lub bezpośrednio mają wpływ na rozstrzygnięcie spraw ludzkich, są bowiem powszechne, masowe, a przy tym niejawne, niezrozumiałe i niekontrolowalne przez adresatów takich decyzji. Również decydenci nie mają bezpośredniego i wyłącznego wpływu na sposób uczenia się algorytmów i podejmowania przez nie decyzji.

[535] O.H. Gandy, *op. cit*.
[536] S. Barocas, A.D. Selbst, *op. cit.*, s. 677.
[537] *Ibidem*.

160:75847

W związku z tym w niniejszym podrozdziale podejmę próbę skonstruowania pojęcia dyskryminacji automatycznej na podstawie przesłanek norm prawa antydyskryminacyjnego. O ile bowiem dyskryminacja automatyczna mieści się w pojęciu analitycznym dyskryminacji i stanowi nową jej formę, o tyle mechanizmy różnicowania prowadzące do wykluczenia algorytmicznego (ang. *algorithmic bias*) będą objęte zakresem istniejącej ochrony antydyskryminacyjnej. Dopiero bowiem zakwalifikowanie dyskryminacji automatycznej jako typu dyskryminacji zakazanej prawem europejskim umożliwi poszukiwanie ochrony w przepisach antydyskryminacyjnych lub w praktyce stosowania tego prawa. Jeżeli zaś definicja analityczna nie będzie mieścić w sobie dyskryminacji automatycznej, będzie trzeba rozważyć jej zmianę z perspektywy celu regulacji.

Wprowadzanie zakazu dyskryminacji przez algorytmy, w tym systemy sztucznej inteligencji, jeżeli procesy krzywdzącego różnicowania wykonywane przez algorytmy nie mieściłyby się w pojęciu dyskryminacji, tworzyłoby jedynie puste normatywnie regulacje. Mechanizm różnicowania algorytmicznego nie byłby bowiem w relewantnej formie bezprawny. Zakres odmienności dyskryminacji automatycznej od dyskryminacji klasycznej będzie zatem podstawą do sformułowania analitycznej definicji tej pierwszej.

Przede wszystkim należy podkreślić, że dyskryminacja automatyczna może być realizowana, wypełniając przesłanki obowiązującego zakazu. Różnicowanie automatyczne bezpośrednio lub pośrednio oparte na cechach chronionych będzie bowiem przypadkiem dyskryminacji, który zgodnie z regulacjami szczególnymi we właściwym trybie będzie podlegać penalizacji. Następnie należy wskazać dwa pozostałe typy dyskryminacji automatycznej, czyli różnicowanie na podstawie kategorii niebędących cechami chronionymi prawem antydyskryminacyjnym oraz różnicowanie, którego procesu i wyniku nie jesteśmy w stanie zdekodować. W dwóch ostatnich grupach zakaz dyskryminacji nie zadziała, gdyż – po pierwsze – taka grupa nie będzie chroniona prawem antydyskryminacyjnym ze względu na brak jej identyfikacji przez chronioną cechę (przy czym kategoria wyróżniająca dla tej grupy będzie zrozumiała dla człowieka); po drugie, tożsamość takiej grupy będzie niezrozumiała ze względu na kategoryzację na podstawie niejasnych lub nieznanych cech i informacji niebędących cechami ludzkimi. W tym kontekście należy odwołać się do przeprowadzonej w rozdziale I analizy pojęcia dyskryminacji.

Jednocześnie aby można było scharakteryzować dany przepis jako normę prawa antydyskryminacyjnego, wyróżnia się m.in.:

1) warunek przesłanek osobistych;
2) warunek grup porównywalnych;
3) warunek praktyki dyskryminacyjnej;
4) warunek względnej krzywdy;
5) warunek nierównej dystrybucji[538].

W tabeli 1 przestawiam porównanie sposobu spełnienia wskazanych warunków.

Tab. 1. Dyskryminacja automatyczna jako odrębne pojęcie prawne

Przesłanka	**Dyskryminacja automatyczna**	**Dyskryminacja w ujęciu prawa europejskiego**
Warunek przesłanek osobistych służących do kategoryzacji	– bezpośrednio oparta na cechach chronionych, które mają wymiar cech osobistych wypływających z godności człowieka; – pośrednio oparta na cechach chronionych; – wysoce intersekcjonalna, krzyżowa i wielokrotna, przez co często oparta na cechach i danych (lub ich zestawach) bezpośrednio skorelowanych z cechami chronionymi (jako substytuty cech chronionych) lub	– bezpośrednio oparta na cechach chronionych, które mają wymiar cech osobistych wypływających z godności człowieka; – pośrednio oparta na cechach chronionych; – cechy służące do kategoryzacji mogą być obserwowane przez ludzi (zarówno autorów, jak i adresatów decyzji); – cechy chronione, ze względu na które następuje różnicowanie, są możliwe do nazwania

[538] Sandra Wachter proponuje weryfikację na podstawie innego doboru cech, które czynią grupy godnymi ochrony, skupiając się na: 1. niezmienności cechy i możliwości jej wyboru; 2. znaczeniu, arbitralności i zasługach; 3. historycznym ucisku, napiętnowaniu i strukturalnym upośledzeniu oraz 4. znaczeniu społecznym grupy. Autorka dochodzi do wniosku, że różnicowanie algorytmiczne w zakresie grup identyfikowanych przez cechy niechronione przez prawo antydyskryminacyjne nie wypełnia przesłanek pojęcia prawnej dyskryminacji; zob. *eadem*, *op. cit.*

162:10576

	pośrednio skorelowanych z cechami chronionymi (jako pochodne danych chronionych); – polegająca również na kategoryzowaniu na podstawie innych cech nieosobistych i danych lub zespołów danych, w tym informacji zaobserwowanych lub danych behawioralnych, niż cechy chronione prawem antydyskryminacyjnym; – cechy i dane służące do kategoryzacji są trudne lub niemożliwe do zaobserwowania w procesie różnicowania (zarówno przez autorów, jak i adresatów decyzji), przez co są niełatwe do skatalogowania i jednoznacznego wskazania	
Warunek grup porównywalnych	– wynikająca ze stosunków dominacji; – niewidoczna, niejawna, nieprzejrzysta zarówno dla adresata decyzji,	– wynikająca ze stosunków dominacji; – widoczna lub możliwa do zaobserwowania dla adresata, decydenta, władzy i społeczeństwa;

163:29706

	decydenta, jak i społeczeństwa; – oparta na cechach chronionych lub na cechach i danych niechronionych prawem antydyskryminacyjnym, przy czym wynik kategoryzacji może uniemożliwiać wskazanie cechy lub informacji, która doprowadziła do zaliczenia pokrzywdzonego do konkretnej grupy, w związku z czym wskazanie grupy porównywalnej nie może zostać przeprowadzone przez eliminację cechy lub danej służącej do kategoryzacji; – wysoce intersekcjonalna, krzyżowa i wielokrotna, przez co ustalenie grupy porównywalnej jest trudne, a nawet niemożliwe; – w jej wypadku osoba dyskryminowana nie jest	– cechy chronione są dostrzegalne/obserwowalne nawet pośrednio, a decydent ma wpływ na określenie cech, dzięki czemu wskazanie grupy, do której przynależy pokrzywdzony, odbywa się na podstawie cech służących do kategoryzacji, a grupy porównywanej w związku z niewystępowaniem tych cech; – w tym wypadku pokrzywdzony ma możliwość ustalenia i utożsamiania się z grupą porównywalną przy zaistnieniu dyskryminacji pośredniej, nawet w przypadku dyskryminacji krzyżowej; – dotyka podmioty z grup charakteryzowanych cechami chronionymi prawem antydyskryminacyjnym

164:12483

	egzemplarycznym przedstawicielem grupy, nie utożsamia się z nią i nie może intuicyjnie podać porównywalnej grupy; – dotyka głównie grup historycznie marginalizowanych, ale również konsumentów lub wyborców (w wyniku zaburzenia autonomii informacyjnej słabszej strony relacji prawnej), jak również innych nieistotnych grup społecznych (np. posiadaczy kart kredytowych)	
Warunek praktyki dyskryminacyjnej	– niewidoczna, niejawna, nieprzejrzysta (efekt czarnej skrzynki powoduje niemożliwość obiektywnego wyjaśnienia procesu i przyczyn uzyskania konkretnego wyniku); – ujawniana przez dyskryminacyjne skutki, rzadziej dyskryminujące traktowanie; – może polegać na klasycznej nierównej	– ujawniana w dyskryminującym traktowaniu; – w przypadku dyskryminacji pośredniej punkt ciężkości przeniesiony z traktowania na skutki

165:82756

	dystrybucji lub na ograniczeniu autonomii informacyjnej prowadzącej do nierównych szans czy możliwości	
Warunek względnej krzywdy	– podmiot dyskryminowany zdaje sobie sprawę z mniej korzystnego położenia z powodu wyniku procesu; – podmiot dyskryminowany może również nie zdawać sobie sprawy z mniej korzystnego położenia na skutek zastosowania różnicowania automatycznego; – nie występuje zaistnienie krzywdy rozumianej wprost w przypadku ograniczenia autonomii informacyjnej; – dyskryminacja automatyczna jest uzasadniana względami statystycznej poprawności, społecznej przydatności, ekonomicznej racjonalności i	– podmiot dyskryminowany zdaje sobie sprawę z mniej korzystnego położenia z powodu widoczności procesów dyskryminacji i ma możliwość rozpoznania swojej sytuacji jako krzywdzącej na skutek procesu różnicowania i/lub jego wyniku

166:10646

	technologicznej neutralności decyzji ferowanych przez ASD, przez co różnicowanie mimo krzywdzących rezultatów w odbiorze społecznym może uchodzić za usprawiedliwione i racjonalne	
Warunek nierównej dystrybucji	– niewidoczna, niejawna, nieprzejrzysta (efekt czarnej skrzynki powoduje niemożliwość obiektywnego wyjaśnienia procesu), a przy tym masowa i intersekcjonalna; – nie ma możliwości wykazania grupy porównywalnej definiowanej przez zbiory danych (również innych niż cechy chronione), co uniemożliwia wykazanie nierównego traktowania	– podmiot dyskryminowany zdaje sobie sprawę z nierównej dystrybucji w toku procesu i/lub na skutek wydanego rozstrzygnięcia

Źródło: opracowanie własne.

Podsumowując, należy wskazać, że dyskryminacja automatyczna to specyficzny rodzaj społecznie szkodliwego i niesprawiedliwego, ekonomicznie napędzanego, masowego, niejawnego i nieprzejrzystego różnicowania, które odbywa się dzięki narzędziom algorytmicznym, w tym systemom sztucznej inteligencji, prowadzącym do kategoryzowania

167:68768

ludzi na podstawie ich cech osobistych chronionych historycznie bądź innych kategorii, w tym danych behawioralnych, w toku którego dochodzi do nierównej dystrybucji dóbr lub ograniczenia autonomii informacyjnej i decyzyjnej jednostki w stosunku do grupy niepodlegającej decyzjom automatycznym. Niekorzystna sytuacja pokrzywdzonych procesem dyskryminacyjnym nie musi mieć przy tym podłoża historycznego ucisku i podporządkowania konkretnej grupy, aczkolwiek wpływ automatycznej dyskryminacji na grupy marginalizowane jest krytyczny pod względem utrzymania tkanki społecznej. Proces algorytmicznego różnicowania jest jednocześnie specyficzny dla kontekstu i celu przetwarzania danych w systemie. Dyskryminacja automatyczna może być też zindywidualizowana i wieloaspektowa, podobnie do dyskryminacji krzyżowej oraz intersekcjonalnej.

Dokonana próba zdefiniowania pojęcia dyskryminacji automatycznej na podstawie warunków analitycznej definicji dyskryminacji unaocznia deficyt teoretycznej refleksji i pogłębionej dyskusji w zakresie celów i pożądanych kierunków rozwoju prawa antydyskryminacyjnego wobec rozwoju automatyzacji procesów decyzyjnych[539]. Algorytmiczne różnicowanie w swoich specyficznych formach wykracza bowiem poza aktualne dopuszczalne rozumienie przedmiotu zakazu dyskryminacji.

W pozostałym zakresie dyskryminacja automatyczna może być kwalifikowana jako dyskryminacja pośrednia. To zaś skutkuje problemami na poziomie ewentualnych postępowań sądowych. Dyskryminacja automatyczna jest bowiem trudna do wykrycia, a ustalenie grupy porównywalnej może się okazać niezwykle problematyczne, a czasami wręcz niemożliwe. Drugą problematyczną kwestią jest to, że zakaz dyskryminacji pośredniej zawiera wyjątek, który umożliwia niekorzystne różnicowanie członków grup chronionych, o ile taki proces może być obiektywnie uzasadniony.

Dyskryminacja automatyczna może się przy tym cechować racjonalnością wynikającą ze statystycznej poprawności, przez co będzie dostarczać pozornie istotnych argumentów za dopuszczalnością różnicowania w toku testu proporcjonalności. Sądy będą zatem zmuszone dokonywać oceny, na ile sytuacja ofiar dyskryminacji byłaby inna w przypadku rezygnacji z modelu algorytmicznego i powiązanej z nim racjonalności ograniczenia kosztów oraz dokładności przewidywania i mitygacji potencjalnych błędów i czy mogłaby mieć miejsce w alternatywnych modelach lub systemach decyzyjnych[540].

[539] M. Domańska, *op. cit.*, rozdz. 1.4.1.
[540] D. Morondo Taramundi, *op. cit.*, s. 80.

168:78646

Solon Barocas proponuje przy tym, aby ewentualne badanie takiego procesu różnicowania, prowadzącego pośrednio do dyskryminujących skutków opierało się na określeniu, czy błędy wydają się możliwe do uniknięcia, ponieważ uzyskanie dostępu do dodatkowych lub bardziej precyzyjnych danych byłoby trywialne bądź nie wiązałoby się z kosztami, które przewyższają korzyści[541]. Taki schemat jest wartościowy, jednak zakłada, że sądy powszechne będą miały dostęp do specjalistycznej wiedzy, pozwalającej im ocenić zarówno zastosowany model, jak i ten porównywalny. Z perspektywy realiów procesu sądowego wymogi te wydają się jednak aktualnie nierealne do osiągnięcia. Skoncentrowanie dyskusji o dyskryminacji automatycznej wokół jej powiązania z dyskryminacją pośrednią skazuje więc podmioty poszukujące ochrony na dużo bardziej rygorystyczne przesłanki procesowe.

Na tym tle uważam, że warto powrócić do analizy teoretycznej warunków prowadzących do zakwalifikowania dyskryminacji automatycznej jako dyskryminacji bezpośredniej. Jak dowiodłam w przeprowadzonym badaniu problemu dyskryminacji automatycznej, zjawisko to w sposób widoczny uderza w grupy społeczne uznane za historycznie marginalizowane i narażone na stygmatyzację i wykluczenie, szczególnie osoby o niskim statusie majątkowym, pochodzące z mniejszości etnicznych, zamieszkujące dzielnice ubogie lub o wysokiej przestępczości, a także osoby o utrudnionym dostępie do edukacji, rynku pracy czy usług socjalnych i opieki zdrowotnej. Osoby zakwalifikowane do tych grup społecznych historycznie były ofiarami dyskryminacji ze względu na różnicowanie na podstawie cech chronionych. Podobną obserwację poczyniła lady Hale w sprawie *Lee v. Ashers*[542], wskazując przy tym, że dyskryminacja bezpośrednia może mieć przez to miejsce również w przypadku dyskryminacji automatycznej, gdy „kryterium wykorzystywane jako powód mniej korzystnego traktowania nie jest samą cechą chronioną, ale jakimś jej substytutem lub korelatem"[543].

Tym samym, rozwijając pogląd sędziny, można by rozważyć na gruncie prawa europejskiego, czy ASD wypełniają warunki dyskryminacji bezpośredniej w przypadku: 1. zakodowania przez człowieka w algorytmie kryterium bezpośrednio powiązanego z cechą chronioną (np. informacji, że związek partnerski to związek różnopłciowy) lub 2. nauczenia się przez ASD doskonałej korelacji między zestawem danych analizowanych przez algorytm a cechą chronioną (np. zestawu danych i cech wskazujących z wielkim prawdopodobieństwem na rasę

[541] S. Barocas, *op. cit.*, s. 719.
[542] Wyrok Sądu Najwyższego Wielkiej Brytanii z 10.10.2018 r. w sprawie *Lee v. Ashers Baking Co*, sygn. UKSC 2017/0020.
[543] *Ibidem*.

169:38601

lub płeć adresata decyzji)[544]. Jeżeli algorytm identyfikuje zestaw danych, które zastosowane razem są rewelacyjnym przybliżeniem chronionej cechy, wówczas może dojść do bezpośredniej dyskryminacji, nawet jeśli z perspektywy ludzkiej rekonstrukcja tych danych byłaby niemożliwa, a człowiek nie zdawałby sobie sprawy z tego, że taki wynik będzie miał miejsce[545].

Rozważenia wymaga przy tym kwestia, czy w takim przypadku kryterium musi być nierozerwalnie związane z cechą chronioną, a jeżeli nie, to z jakim prawdopodobieństwem musi prowadzić do tej cechy. Jeremias Adams-Prassl proponuje, aby przyjmować istnienie działania na podstawie substytutu cechy chronionej, o ile: 1. algorytm daje korzystniejsze wyniki w wypadku podobnie usytuowanej osoby, która nie dysponuje chronioną cechą (np. przy identycznym podaniu o pracę, ale z danymi męskimi, a nie kobiecymi); oraz 2. wyniki algorytmu są konsekwentnie stronnicze na korzyść danej cechy chronionej (np. kobiety otrzymują średnio niższe wyniki niż mężczyźni)[546]. Warunki te uzupełniłabym, zgodnie z koncepcją F.K. Thomsena, o istotność grupy społecznej, której dotyka proces dyskryminacji automatycznej. Jeżeli bowiem różnicowanie wpływa na sytuację grupy politycznie niedoreprezentowanej, społecznie wykluczonej czy w złej sytuacji materialnej, co historycznie doprowadziło do marginalizacji, podporządkowania i jej zniewolenia, to uwaga związana z różnicowaniem jest kluczowa. Uważam przy tym, że zaprezentowany kierunek rozważań nad formą bezpośrednią dyskryminacji automatycznej jest niezwykle cenny i wart pogłębienia, chociaż oczywiście wymagałby dokonania istotnego zwrotu w sposobie stosowania przepisów antydyskryminacyjnych przez sądy. Prezentowany kierunek zdaje się też odpowiadać kierunkowi orzeczniczemu TSUE w sprawie *Coleman v. Attridge Law i Steve Law*[547], w wyroku której Trybunał dopuścił dyskryminację przez asocjację rodzica z osobą z niepełnosprawnością, dzięki czemu otworzył drogę do uznania za dyskryminację bezpośrednią niektórych przypadków dyskryminacji przez korelację i błędną kategoryzację.

Należy jeszcze wyraźnie podkreślić, że wraz z ewolucją stosunków społeczno-polityczno-gospodarczych prawo musi dokonywać reinterpretacji języka ustawowego. Zakaz dyskryminacji ma na celu ograniczanie rzeczywistego problemu dominacji, w związku z czym

[544] J. Adams-Prassl, R. Binns, A. Kelly-Lyth, *op. cit.*, s. 160.
[545] *Ibidem*.
[546] *Ibidem*, s. 23.
[547] Wyrok TSUE z 17.07.2008 r. w sprawie *S. Coleman v. Attridge Law*…

170:89639

musi kroczyć za zjawiskami decentralizacji władzy i wpływu tych procesów na brak odpowiedniej ochrony praw i wolności jednostki.

Pamiętajmy, że automatyczna dyskryminacja może przybrać formę dyskryminacji zarówno pośredniej, jak i bezpośredniej. Obie formy mieszczą się w języku ustawowym i ustalonych granicach normatywnych, ale nie mieszczą się wprost w kategoriach zdefiniowanych przez sąd[548]. Reinterpretowanie wymogów ustawowych stawianych dyskryminacji bezpośredniej w kierunku specyficznym dla ASD zamknęłoby pole do racjonalizowania i statystycznego uzasadniania krzywdzącego działania algorytmu. Ochroną objęte zostałyby przypadki różnicowania na podstawie wysoce intersekcjonalnych kategorii, albowiem identyfikacja grupy defaworyzowanej i faworyzowanej nie byłaby tak trudna[549].O ile bowiem dyskryminacja automatyczna mogłaby w większej liczbie przypadków zostać uznana za bezpośrednią, o tyle ochrona na gruncie obecnych przepisów nie doznawałaby wielu ograniczeń związanych z procesowym uprawdopodobnieniem przesłanek jej zaistnienia. Ciężar dowodowy algorytmów zaliczanych do czarnoskrzynkowych mógłby w takich przypadkach obciążać operatorów ASD, a ci byliby zobowiązani do uwzględniania nie tylko standardów antydyskryminacyjnych w fazie projektowania rozwiązania algorytmicznego, ale też oceny skutków stosowania sytemu na zasadach obecnie realizowanych w ramach regulacji RODO.

Zaprezentowana próba redefiniowania zakresu pojęcia dyskryminacji nie rozwiązuje jednak w pełni problemu różnicowania z powodu zestawów cech chronionych, jeśli kategorie te jedynie z pewnym prawdopodobieństwem (a nie na skutek doskonałej korelacji) łączyłyby się z cechą chronioną, a mimo wszystko prowadziłyby do krzywdzącego różnicowania. Jak wskazują Batya Friedman i Helen Nissenbaum[550], elementem koniecznym definicji dyskryminacji jest konkretny katalog podstaw zróżnicowanego traktowania. Różnicowanie nie musi też być niesprawiedliwe (w sensie moralnym) ani wypełniać wszelkich przesłanek dyskryminacji (w sensie prawnym).

Na gruncie tej myśli warto odwołać się do pojęcia dyskryminacji intersekcjonalnej, która również nie daje się łatwo odwzorować w kategoryzowaniu – ze względu na ugruntowane chronione prawem antydyskryminacyjnym cechy dopuszcza bowiem wielorodzajowe różnicowanie[551]. Ta forma dyskryminacji, dostrzeżona na skutek obserwacji sytuacji

[548] J. Adams-Prassl, R. Binns, A. Kelly-Lyth, *op. cit.*
[549] D. Morondo Taramundi, *op. cit.*, s. 79.
[550] B. Friedman, H. Nissenbaum, *Bias in Computer Systems*, „ACM Transactions on Information Systems" 14, 1996, nr 3, s. 330–347.
[551] M. Mann, T. Matzner, *op. cit*.

171:69123

afroamerykańskich kobiet, których krzywda – jako ofiar dyskryminacji zarówno ze względu na rasę, jak i płeć – była znacząco większa niż grup dyskryminowanych z uwagi na jedną z tych cech, jest niezwykle istotna z perspektywy poszukiwania ochrony przed dyskryminacją automatyczną. Dyskryminacja intersekcjonalna pozwala bowiem na uwzględnienie wyniku marginalizacji na skutek połączenia wielu cech, które nie muszą być redukowane do jednej kategorii i jednorodzajowej grupy chronionej. To zaś pozwala uznać, że analiza algorytmiczna prowadząca do wyniku dyskryminującego na podstawie zestawu danych i informacji przypomina mechanizmy leżące u podstaw dyskryminacji intersekcjonalnej[552]. Co więcej, formy dyskryminacji intersekcjonalnej są trudne do udowodnienia z perspektywy zidentyfikowania i porównania grup faworyzowanych z defaworyzowanymi. Nawet jeśli różnicowanie dotyczy rasy i płci, to dane tyczące się grup porównywalnych mogą być niedostępne, natomiast w przypadku innych kategorii populacje mogą być statystycznie zbyt małe[553].

Z tej perspektywy możliwym podejściem do problemu dyskryminacji automatycznej byłoby stypizowanie grup marginalizowanych na gruncie dyskryminacji intersekcjonalnej, a następnie zastosowanie do nich klasycznych instrumentów prawa antydyskryminacyjnego[554]. W ostatnich latach opracowano formalne metody, które pozwalają odkryć na podstawie danych, że pewna grupa (być może wcześniej nieznana w tej kombinacji) jest objęta segregacją[555] lub dyskryminowana[556]. Te narzędzia mogłyby się przyczynić do precyzyjnego nazwania nowej cechy, która powinna uzupełnić katalog cech chronionych europejskim prawem antydyskryminacyjnym.

Zastosowanie teorii intersekcjonalnej wzbudza jednak pewne wątpliwości. Dyskryminacja intersekcjonalna orientuje się wszak wokół cech, które zostały już uznane za źródło dyskryminacji. W tym zakresie z pomocą może przyjść przedstawiona wcześniej teoria ścisłej

[552] *Ibidem*.
[553] K. Crenshaw, *Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics*, „The University of Chicago Legal Forum" 1, 1989, s. 146.
[554] M. Mann, T. Matzner, *op. cit*.
[555] Celem narzędzia SCube jest wspieranie wykrywania segregacji z danych relacyjnych i grafowych. System ten opiera się na przypisanym grupowaniu grafów i eksploracji zestawów danych, oferuje analitykowi wielowymiarową kostkę segregacji danych do eksploracyjnej analizy danych; por. A. Baroni, S. Ruggieri, *SCube: A Tool for Segregation Discovery*, [w:] *Advances in Database Technology – 22nd International Conference on Extending Database Technology*, EDBT, Lisbon 2019, s. 542–545, https://openproceedings.org/2019/conf/edbt/EDBT19_paper_212.pdf (dostęp: 4.12.2021).
[556] M. Kearns *et al., Preventing Fairness Gerrymandering: Auditing and Learning for Subgroup Fairness. Proceedings of the 35th International Conference on Machine Learning*, „Proceedings of Machine Learning Research" 80, 2018, s. 2564–2572, https://proceedings.mlr.press/v80/kearns18a.html (dostęp: 4.12.2021).

172:11572

korelacji zestawu danych z cechą chronioną. Pojawiające się formy dyskryminacji automatycznej mogą jednak prowadzić do niesprawiedliwego różnicowania również na podstawie danych, które same w sobie wydają się neutralne lub są danymi zaobserwowanymi (np. dane behawioralne).

Uzupełniając te zastrzeżenia, Matthias Leese wskazuje, że mogą się też pojawić zupełnie nowe formy dyskryminacji, które nie mają charakteru reprezentatywnego[557]. W swojej analizie również dostrzegam ten rodzaj algorytmicznego różnicowania, na skutek którego tożsamość grupy defaworyzowanej nie będzie niezrozumiała ze względu na dokonaną kategoryzację, jako że samo różnicowanie będzie oparte na niezrozumiałych lub nieznanych cechach i informacjach niekoniecznie będących cechami osobistymi. Z tej przyczyny nowe grupy dyskryminowane mogą być tworzone przez kombinacje cech wejściowych, które nie mają nawet sensu jako intersekcjonalna kombinacja tożsamości[558]. Oznacza to, że dane zarówno wejściowe, jak i wyjściowe systemów algorytmicznych mogą nie mieć bezpośredniego związku z cechą klasycznie chronioną prawem antydyskryminacyjnym, a nadal może stać się tak, że system algorytmiczny systematycznie będzie stawiał w niekorzystnej sytuacji osoby mające, powiedzmy, określoną kombinację historii przeglądania, marki posiadanego komputera i ulubionych zespołów, wynikających np. z polubień na Facebooku[559].

Można przy tym rozważyć, czy pozostawienie tych grup, które nie są pod względem oceny moralnej różnicowania podobne do tradycyjnych grup chronionych, będzie naruszać teoretyczne podstawy, cel i zadania prawa antydyskryminacyjnego. Prawo antydyskryminacyjne nie inkorporuje bowiem obowiązku racjonalnego podejmowania decyzji przez decydentów, a zatem nie można go identyfikować z zakazem nieprzewidywalnych czy nieprzejrzystych decyzji[560].

W mojej ocenie przeciwdziałanie dyskryminacji automatycznej powinno się koncentrować na ochronie istotnych grup społecznych, a więc tych, które historycznie doświadczały wykluczenia i marginalizacji. Grupy te są politycznie niewidoczne i niereprezentowane proporcjonalnie do skali zagrożenia w procesie stanowienia prawa i będą nadal w nierównym tempie uzyskiwały odpowiednią reprezentację[561]. Brak właściwej reprezentacji połączony z trudniejszą sytuacją finansowo-socjalną utrudnia zaś walkę z nierównościami i wykluczeniem.

[557] M. Leese, *op. cit*.
[558] M. Mann, T. Matzner, *op. cit.*
[559] *Ibidem*, s. 5.
[560] S. Wachter, *op. cit.*, s. 41.
[561] S. Barocas, A.D. Selbst, *op. cit.*, s. 688.

173:76242

Niedostateczny potencjał samodzielnego przeciwdziałania niesprawiedliwemu różnicowaniu powoduje z kolei, że w tych grupach procesy pogłębiające wykluczenie są najszybsze i trwale naruszają tkankę społeczną. Z tej przyczyny przyjmuję, że ochrona grup uciśnionych w toku historycznych procesów dominacji jest kluczowa z perspektywy adaptacyjnej roli prawa antydyskryminacyjnego, którego celem jest zatrzymanie niedopuszczalnych skutków społecznych krzywdzącego różnicowania. W przypadku różnicowania prowadzącego do pokrzywdzenia grup niereprezentatywnych poszukiwałabym ochrony prawnej na gruncie przepisów ochrony danych osobowych i prawa do prywatności.

W niniejszym podrozdziale przeprowadziłam analizę pojęcia dyskryminacji automatycznej z uwzględnieniem terminu dyskryminacji przedstawionego w rozdziale I, a także z perspektywy możliwości adaptacji formy różnicowania algorytmicznego do istniejących ram prawnych i oczekiwanych kierunków adaptacyjnego rozwoju prawa antydyskryminacyjnego. Dyskusja ta jest potrzebna i doniosła badawczo nie tylko ze względu na skuteczne przeciwdziałanie dyskryminacji automatycznej, ale też konieczność pogłębionej refleksji nad celem nowożytnego prawa dyskryminacyjnego i dalszym kierunkiem urzeczywistniania gwarancji równościowych. Dyskryminacja związana z eksploracją danych wymusi bowiem konfrontację pomiędzy dwoma rozbieżnymi zasadami leżącymi u podstaw prawa antydyskryminacyjnego: równością w ujęciu formalnym i materialnym.

2. Ryzyko jako podstawa regulacji

Algorytmy mogą różnicować zarówno podobnie do ludzi – powielając uprzedzenia historyczne zapisane w danych – jak i tworząc nowe typy grup społecznych. Ta nowość może mieć przy tym dwa oblicza. Pierwsze z nich polega na tworzeniu grup poprzez identyfikację ludzi na podstawie cechy nieuznanej dotychczas za chronioną prawem antydyskryminacyjnym (np. posiadanie dzieci, miejsce zamieszkania). Drugi zaś opiera się na kategoryzowaniu na bazie niezrozumiałych dla człowieka wyników analizy algorytmicznej (np. zaszeregowanie przez algorytm użytkowników telefonów z niewiadomych powodów do konkretnych zespołów – w toku eksperymentu naukowego opisanego w rozdziale II)[562]. W tym drugim przypadku nie chodzi o to, aby sama cecha była niezrozumiała dla człowieka (np. metadane), może być ona bowiem pojmowalna (np. posiadanie psa), ale o to, że na skutek zaprezentowanego przez

[562] Podobnie: S. Wachter, *op. cit*.

174:10860

algorytm wyniku procesu decyzyjnego człowiek nie jest w stanie odwzorować relewantnej dla procesu różnicowania informacji. Tak utworzone grupy może łącznie cechować heterogeniczny lub monolityczny dobór kryteriów różnicowania, społeczna doniosłość cech lub jej brak, a nadto dokładność klasyfikacji lub jej brak[563].

Jednocześnie procesy różnicowania prowadzące do takiej wielopłaszczyznowej i skomplikowanej kategoryzacji mogą być moralnie karygodne z uwagi na potęgowanie wykluczenia i upośledzenia grup historycznie doświadczających znacznej niesprawiedliwości w społeczeństwie. Ten historyczny komponent prowadzi do wniosku, że nadal aktualnym celem zakazu dyskryminacji może być ograniczenie realnego problemu dominacji i systemowej opresji. Zmianie uległ jednak rzeczywisty obraz tej supremacji. Zagrożenia tradycyjnie związane z problemem dyskryminacji miały wszak zmysłowy charakter i można je było zaobserwować. Widoczna była również ich skala oraz następująca w ich skutku walka o równość, tocząca się między członkami społeczeństwa a władzą mającą zdolność do rządzenia nim. Różnicowanie opierało się zaś na cechach umożliwiających identyfikację grup nierówno traktowanych, co pozwalało też na wykazanie lub uprawdopodobnienie dyskryminującego procederu. Efekty kategoryzowania ludzi według bezprawnych założeń były dostrzegalne nie tylko przez podmioty dotknięte dyskryminacją, ale również przez pozostałe grupy społeczne. To z kolei umożliwiało sprzeciw wobec niesprawiedliwego różnicowania i czyniło podwaliny pod poszerzanie pojęcia dyskryminacji o różnicowanie oparte bezpośrednio lub pośrednio na kolejnych, nowych cechach chronionych.

Prawo dyskryminacyjne wspierające walkę o równość stanowiło wówczas ograniczenie dla samego dyskryminatora, będącego elementem aparatu państwowego bądź dominatora. Obecnie zmienia się nie tylko proces różnicowania, ale też jego aktorzy. Konsekwencją emergentności dyskryminacji automatycznej, niewidoczności jej kryteriów i niejawności mechanizmów jest inherentne ryzyko zaistnienia nieuzasadnionego różnicowania w toku wykorzystywania automatycznych systemów decyzyjnych. Nadto domyślnym oprawcą przestaje być rządzący, a stają się nim nowi aktorzy, którzy otrzymują władzę różnicowania (np. Google, Facebook, Amazon czy Alibaba).

Empiryczne dowody na stronniczość ASD uzasadniają ponowne przeanalizowanie, a być może nawet redefiniowanie, wielu podstawowych zagadnień dla konceptualnego aparatu prawa antydyskryminacyjnego, począwszy od doktrynalnych debat dotyczących normatywnych

[563] *Ibidem*, s. 13.

175:40159

podstaw tej dyscypliny, na rekonstrukcji podstawowych definicji skończywszy, a ostatecznie reformę sądowej kontroli realizacji zakazu dyskryminacji, w tym zmianę reguł ciężaru dowodowego[564]. W większości przypadków automatyczne systemy decyzyjne nie będą bowiem różnicowały w sposób jawny i zrozumiały dla człowieka, nie będzie to także powiązane z fizycznymi aktami dostrzeganymi w przestrzeni publicznej. Niewidoczność dyskryminacji automatycznej, oderwanie różnicowania od cech historycznie wykształconych jako społecznie doniosłe dla tego procesu oraz nieprzejrzystość ASD to poważne utrudnienie przy realizacji ochrony na podstawie istniejącego prawa antydyskryminacyjnego, szczególnie z powodu niewystarczających kompetencji poznawczych po stronie organów nadzoru. Nie ma również innych regulacji, które w sposób komplementarny do prawa antydyskryminacyjnego objęłyby penalizacją różnicowanie automatyczne.

Skoro nie można zakładać, że algorytmy będą dyskryminować według cech obserwowalnych lub zrozumiałych dla człowieka[565], a niekorzystne traktowanie użytkowników w sposób pośredni i nieoczywisty może się wymykać tradycyjnym definicjom dyskryminacji[566], należy przyjąć, że w przypadku decyzji ASD zawsze istnieje ryzyko naruszeń praw i wolności na skutek niesprawiedliwego różnicowania. Przeciwdziałanie zjawisku niesprawiedliwego algorytmicznego różnicowania nie będzie przy tym mogło zostać ograniczone poprzez klasyczne formy scentralizowanego nadzoru i władzy publicznej. Dlatego też wobec stałego ryzyka dyskryminującego skutku stosowania automatycznych systemów decyzyjnych, powiązanego ze zmianą aktorów procesu dominacji, powstaje konieczność analizy i konkretyzacji metod stosowanej oceny tego ryzyka i poszerzenia świadomości organizacji prywatnych o potrzebę wdrożenia wewnętrznych procedur monitorujących ASD.

Implementacja podejścia opartego na ryzyku (ang. *risk-based approach*) prowadzi do rozszerzenia odpowiedzialności ponad zasady ogólne wymagające zaistnienia szkody, ale i budowania świadomości ryzyka dyskryminacji. System oceny ryzyka jako narzędzie kontroli zakłada przy tym strukturę pionową, doprowadzającą do monitorowania ryzyka od niższych szczebli do organów państwowych. Zarządzanie ryzykiem powinno jednocześnie prowadzić do przestrzegania obowiązków przez adresatów regulacji i zdefiniowania warunków, których

[564] J. Adams-Prassl, R. Binns, A. Kelly-Lyth, *op. cit.*, s. 175.
[565] B. Mittelstadt *et al.*, *op. cit.*
[566] Ch. Sandvig *et al.*, *Auditing Algorithms : Research Methods for Detecting Discrimination on Internet Platforms*, [w:] *Data and Discrimination: Converting Critical Concerns into Productive: A Preconference at the 64th Annual Meeting of the International Communication Association*, Seattle 2014, https://www.kevinhamilton.org/share/papers/Auditing%20Algorithms%20--%20Sandvig%20--%20ICA%202014%20Data%20and%20Discrimination%20Preconference.pdf (dostęp: 4.12.2021).

176:55955

powinni przestrzegać. Kwintesencją odmienności zarządzania ryzykiem w wypadku ASD jest zatem zmiana oczekiwań wobec rządów, polegająca na zwiększeniu nacisku na bezpieczeństwo i wprowadzeniu zdecentralizowanych form wykonywania władzy[567].

Podejście oparte na ryzyku inspiruje aktualnie tworzone prawodawstwo. Jest ono szczególnie widoczne na poziomie legislacji unijnej, m.in. w RODO i projektowanym AIA, jednak w mojej ocenie bez upowszechnienia tej koncepcji w strukturach głównych aktorów obecnego procesu, w tym u gigantów informatycznych, nie dojdzie do zrealizowania jej celów. Założenia AIA są już szeroko komentowane z uwagi na ukształtowanie obowiązków ze względu na zaszeregowanie rozwiązania do poziomu ryzyka – projekt przewiduje, że systemy niedopuszczalnego ryzyka są wyłączone ze stosowania, zaś SI sklasyfikowana na poziomie ryzyka wysokiego uzależniona jest od spełnienia licznych wymogów, w tym stałego monitorowania ryzyka w systemie zarządzania oraz stosowania procedur testowania[568]. Jednocześnie systemy sztucznej inteligencji wysokiego ryzyka spełniające normy zharmonizowane lub części tych norm, do których odniesienia opublikowano w Dzienniku Urzędowym Unii Europejskiej, uznaje się za spełniające wymogi rozporządzenia[569].

„Myślenie ryzykiem" w AIA jest zatem w pewnym zakresie powieleniem rozwiązań już istniejących. Kod algorytmu powinien być wszak niedyskryminujący analogicznie do przewidzianej w RODO zasady ochrony prywatności na etapie projektowania (ang. *privacy by design*), a monitorowaniem powinien być objęty, podobnie jak w przypadku rozwiązań stosowanych na gruncie RODO, każdy rodzaj potencjalnego, jak również zaistniałego niepożądanego skutku, ocenianego na podstawie bardzo szerokiej skali, począwszy od wpływu na konkretną osobę, której dotyczy proces decyzyjny, aż po oddziaływanie ogólnospołeczne (np. utrata zaufania społecznego)[570]. W tym zakresie obowiązki, analogicznie do RODO, obciążałyby każdy podmiot wprowadzający lub korzystający z automatycznych systemów decyzyjnych, obligując go do stosowania odpowiednich środków technicznych i organizacyjnych, aby móc wykazać zgodność systemu z prawem. Zgodnie z podejściem opartym na ryzyku nie wydaje się przy tym właściwe wprowadzanie w AIA zwolnień z poszczególnych obowiązków jedynie na podstawie tego, czy automatyczny system decyzyjny

[567] J.T. Hryniewicz, *Społeczeństwo ryzyka. Teoria, model, analiza krytyczna*, „Przegląd Socjologiczny" 2, 2014, s. 29.
[568] Art. 9 AIA.
[569] Art. 40 AIA.
[570] Tak: Grupa Robocza Art. 29 w dokumencie 14/EN WP 218.

177:41537

nie został zakwalifikowany do unijnych systemów sztucznej inteligencji wysokiego ryzyka zgodnie z załącznikiem III do rozporządzenia.

Ustanawiane obowiązki muszą być dodatkowo do spełnienia w perspektywie technologicznej. Skonstruowanie obowiązków na zbyt ogólnym poziomie definicyjnym w stosunku do nieprecyzyjnie określonego kręgu podmiotów czy nierealnych w kontekście ograniczeń związanych z aktualnym rozwojem sztucznej inteligencji doprowadzi bowiem do pozorności zarządzania ryzykiem, jego rozliczalności i całego procesu audytowania systemów SI. Jednostki certyfikujące i nadzorujące powinny zatem zostać wyposażone w odpowiednie kompetencje i narzędzia do weryfikacji działania omawianych systemów.

Jednym z zarzutów stawianych AIA są niedoskonałości pojęciowe. Akt posługuje się bowiem definicjami niejasnymi lub niepełnymi, które na dalszym etapie prac mogą jeszcze ulec zmianie. W rozporządzeniu brakuje również definicji ryzyka i systemu zarządzania nim. W tym zakresie można się odwołać do definicji ryzyka przywołanej w rozdziale IV niniejszej pracy. Nowość kontekstu technologicznego dla problemu dyskryminacji uzasadnia także teoretycznoprawną refleksję nad tym pojęciem, przy czym pełna analiza zagadnienia ryzyka wykracza poza ramy przedmiotowe tego opracowania[571]. Należy przy tym za Ulrichem Beckiem[572] zauważyć, że z perspektywy technologicznego kontekstu badania jakościowa odmienność nowego typu ryzyka polega na tym, że przeciętny człowiek nie jest w stanie samodzielnie go oszacować. Ryzyko staje się zatem nieobliczalne i wymyka możliwościom kontroli ludzkiej czy kompensacji. Ofiara dyskryminacji automatycznej nie może bowiem liczyć nawet na pomoc nauki w tym zakresie, jako że procesy głębokiego uczenia maszynowego pozostają niezrozumiałe również dla specjalistów.

Pojęcie inherentnego ryzyka naruszeń należy przy tym rozumieć w tym sensie, że ryzyko nie jest problemem do rozwiązania, a raczej jest powiązane z poszukiwaniem metody kontrolowania i zarządzania przypadkiem przez jego odpowiednie rozpoznanie[573]. Należy dodać, iż zgodnie z ustaleniami Janusza T. Hryniewicza rozwój sił wytwórczych prowadzi do powstania nowego typu zagrożeń, tzw. pozazmysłowych i niepoliczalnych[574], jak np. niejawne i nieprzejrzyste procesy różnicowania. Emergentne zagrożenie ryzykiem o takim charakterze i

[571] Poszerzoną analizę teoretycznoprawną zagadnienia ryzyka przeprowadzili Maciej Pichlak i Klaudia Gaczoł, *op. cit*.
[572] U. Beck, *Społeczeństwo ryzyka. W drodze do innej nowoczesności*, przeł. S. Cieśla, Warszawa 2002.
[573] M. Pichlak, K. Gaczoł, *op. cit.*, s. 38.
[574] J.T. Hryniewicz, *op. cit.*, s. 13.

178:57966

konieczność stałego zarządzania nim wpływa zaś na powstanie społeczeństwa ryzyka[575], które różni się od społeczeństwa przemysłowego rozmiarem i zasięgiem ryzyka tworzonego przez nowe technologie i procesy racjonalizacji oraz musi stosować nowe metody obchodzenia się z wymykającym się możliwościom kontroli ludzkiej ryzykiem[576]. Konieczne jest z tego powodu rozważenie pozaregulacyjnych mechanizmów, które mogłyby zrekompensować dyskryminujące efekty automatyzacji procesów decyzyjnych, przez co rozumiem systemy umożliwiające śledzenie efektów sprzężenia zwrotnego (czyli umacniania się algorytmu na skutek powielania błędu w dyskryminujących wynikach), a także systemy wyrównywania szans w obszarach narażonych na historycznie utrwalone wykluczenie i marginalizację. Wymagany jest przy tym nie tylko nadzór państwowy w formie efektywnego monitoringu instytucjonalnego, ale również nowy ład korporacyjny i wsparcie państwowe dla tego sektora NGO[577]. Wspólną cechą dyskryminacji w rozumieniu prawa europejskiego i dyskryminacji automatycznej jest ich związek ze strukturalnymi wzorcami nierówności, dominacją i hierarchiami społecznymi, które są wbudowane w każdy model społeczny, a także zbiorowy wymiar ciężarów i przywilejów społecznych[578]. Konieczna jest zatem obserwacja tych obszarów i refleksja nad możliwością redukcji nierówności.

3. W kierunku materialnej równości

Niemożliwość efektywnej kontroli organów władzy w domyślnych formułach jej sprawowania oraz nieprzejrzystość, nowość i tempo zmian technologicznych prowadzą do przesunięcia i decentralizacji władzy publicznej. W niektórych przypadkach eksploracja danych sprawi, że skorygowanie dyskryminacyjnych wyników będzie po prostu niemożliwe bez zaangażowania się w kwestię tego, jaki poziom materialnej nierówności jest sprawiedliwy lub dopuszczalny w danym kontekście[579]. Z tej perspektywy – komplementarnie do dyskusji o konieczności poszerzenia zakazu dyskryminacji o jej automatyczne formy – należy rozważyć sposób realizacji postulatów równościowych w nowożytnych demokracjach.

[575] *Ibidem*, s. 17.
[576] A. Cudowska, *Tożsamość społeczna w „innej nowoczesności". Ulricha Becka koncepcja społeczeństwa ryzyka*, [w:] *Etniczność i obywatelskość w nowej Europie. Konteksty edukacji międzykulturowej*, red. J. Nikitorowicz, D. Misiejuk, M. Sobecki, Białystok 2007, s. 95.
[577] M. Mercat-Bruns, *Multiple Discrimination and Intersectionality: Issues of Equality and Liberty*, „International Social Science Journal" 67, 2018, nr 223–224, s. 49.
[578] D. Morondo Taramundi, *op. cit.*, s. 82.
[579] S. Barocas, A.D. Selbst, *op. cit.*, s. 676.

179:10029

Zaobserwowane procesy dyskryminacyjne w ASD – z racji swojej niewidoczności i nieprzejrzystości – z dużym prawdopodobieństwem nie doprowadzą do walki o równość porównywalnej z okolicznościami historycznymi opisywanymi w rozdziale I niniejszej pracy. Brak społecznego oburzenia niesprawiedliwością systemową, a jednocześnie przekonanie o niewielkim zakresie możliwych do wdrożenia rozwiązań determinowanych specyfiką uczenia maszynowego może przy tym pozbawić aktualną dyskusję impetu potrzebnego do głębszych zmian społecznych. Nieodzowna jest przy tym – co już wyjaśniałam – szczególna uwaga władzy publicznej, zorientowana na grupy historycznie marginalizowane i niedoreprezentowane. Osiągnięcie tego efektu będzie tym trudniejsze, że tematyka walki o równość ukazuje konflikt między koniecznością stworzenia przestrzeni do zmian polityczno-społecznych a dążeniem samej władzy do utrzymania kontroli i narzędzi potrzebnych jej do rządzenia w możliwe szerokim zakresie.

Chociaż informatyka sama w sobie nie może rozwiązać problemów społecznych, jej metody mogą być wykorzystywane do diagnozowania i precyzyjnego charakteryzowania tych problemów[580]. Dyskusja o wpływie automatycznych systemów decyzyjnych na prawa człowieka staje się zatem fragmentem opisu rzeczywistości dominacji nad grupami marginalizowanymi, ukazującym wpływ historycznie utrwalanych uprzedzeń na procesy decyzyjne. Systemy automatyczne operujące na dużych zbiorach danych dają bowiem nieograniczone możliwości poszerzenia predykcji i nadzoru, na skutek którego procesy wykluczenia grup społecznych będą się pogłębiać, a schematy wykluczenia – jako statystycznie obserwowalne wzorce zachowań ludzkich – utrwalać. Co więcej, nadzór z powodu decentralizacji władzy przestaje być jedynie narzędziem rządów, stając się źródłem budowania kapitału prywatnego.

Pierwotna logika dyskryminacji orientująca się wokół relacji instytucje państwowe–obywatele zmienia się – pojawiają się w związku z tym próby zdefiniowana tego zjawiska na gruncie polityczno-społecznym. Shoshana Zuboff określiła tę zmianę mianem kapitalizmu inwigilacji[581]. Autorka w książce *Wiek kapitalizmu inwigilacji* stawia tezę, że konkurencja i dynamika rozwoju nowych rynków zmusza „kapitalistów nadzoru”[582] do pozyskiwania

[580] R. Abebe *et al.*, *op. cit.*, s. 253.
[581] S. Zuboff, *Wiek kapitalizmu inwigilacji*, przeł. A. Unterschuetz, Poznań 2020.
[582] Podmioty traktujące informacje zapisane cyfrowo jako darmowy surowiec do produkcji danych behawioralnych, które są następnie kapitalizowane.

180:11264

nadwyżek danych behawioralnych oferujących coraz to większą przewidywalność[583]. Sektor prywatny upatruje zatem szansy w automatyzacji procesów nie tylko ze względu na ograniczenie kosztów prowadzonej działalności operacyjnej, rozumiane jako niezwiększanie zasobów, ale również z uwagi na możliwość realizowania wysoce celowanych kampanii marketingowych. Decydenci technologicznych gigantów, do których władza odpłynęła, mogą być przy tym skłonni tolerować wyższe wskaźniki błędnych decyzji w przypadku niektórych grup, ponieważ uzyskują z tego tytułu wymierne korzyści. Jak bowiem wyjaśnia F. Schauer[584], decydenci, którzy polegają na statystycznie uzasadnionych, ale nieuniwersalnych uogólnieniach, „są jednocześnie racjonalni i niesprawiedliwi". Ta relacja pomiędzy nową zdecentralizowaną formą władzy a człowiekiem jest jednocześnie naznaczona asymetrią informacji oraz nierównowagą wiedzy i siły decyzyjnej, która stawia technologicznych gigantów w faworyzowanej pozycji.

Przejście z wiedzy do władzy jest w ocenie S. Zuboff niezwykle niebezpieczne dla demokracji, ponieważ zautomatyzowane procesy nie tylko gromadzą informacje o zachowaniu ludzi, ale także kształtują je na dużą skalę. Ewolucja procesów decentralizacji połączona z postępem technologicznym doprowadziła wszak do tego, że nowa władza wie wszystko o swoich podwładnych, podczas gdy jej działania są tak zaprojektowane, by pozostawały nieznane dla społeczeństwa.

Omawiane w niniejszej dysertacji przypadki dyskryminacji automatycznej, które wynikały z błędu analizy systemowej bądź biznesowej ASD, tj. błędów ludzkich w projektowaniu architektury danych tych systemów, równie dobrze mogłyby funkcjonować w świecie bez SI oraz udziału komputerów i być przy tym równie krzywdząco dyskryminujące. W wypadku procesu automatycznego mamy jednak do czynienia z obsługą masową, przez co tworzy się efekt skali dyskryminującego algorytmu.

Zadaniem rządów i organizacji międzynarodowych jest nadać kierunek działaniom powszechnie uświadamiającym masowość i skalę zagrożeń związanych z niejawnością oraz wpływem krzywdzących algorytmów na prawa jednostek, w tym przez przywrócenie równowagi informacyjnej.

[583] Nowym sposobem produkcji w kapitalizmie nadzorczym są algorytmy. Prognozy są sprzedawane na rynku, co pozwala zainteresowanym na doskonałą predykcję zachowań użytkownika Internetu.
[584] F. Schauer, *Profiles, Probabilities, and Stereotypes*, London 2003.

181:10369

Sandra Fredman[585] przekonuje, że krzyżujące się relacje władzy można analizować i przeciwdziałać im za pomocą czterech wymiarów:

1) potrzeby zadośćuczynienia niekorzystnej sytuacji;
2) potrzeby zajęcia się stygmatyzacją, stereotypami, uprzedzeniami i przemocą;
3) potrzeby ułatwienia głosu i uczestnictwa;
4) potrzeby dostosowania się do różnicy i zmiany struktur dyskryminacji[586].

Na tej podstawie można uznać, że eksploracja algorytmiczna danych w automatycznych systemach decyzyjnych powinna się koncertować wokół przywrócenia równowagi pomiędzy dwoma uzasadnieniami prawa antydyskryminacyjnego – eliminacją celowej dyskryminacji i wyrównaniem statusu społeczności historycznie pokrzywdzonych[587].

Musimy jednak brać pod uwagę, że nowe formy nadzoru jako nowe technologie władzy nigdy nie zostaną trwale usunięte z bieżącej dynamiki walki politycznej, w związku z czym zmiany polegające na wdrożeniu i rozwinięciu narzędzi do zarządzania ryzykiem dyskryminacji w automatycznych systemach decyzyjnych, dyskusja o możliwych kierunkach redefiniowania procesów dyskryminacji poprzez adaptację do rozwoju technologii oraz wprowadzenie odpowiednich narzędzi umożliwiających egzekwowanie prawa powinny być poparte głębszymi dyskusjami prawnymi w zakresie rozumienia i celów gwarancji równościowych[588]. Poprawa świadomości społecznej co do kształtu i wymiaru nowych form nadzoru umożliwi szerszą dyskusję, która z uwagi na partycypację różnych grup społecznych może wpłynąć na potencjał przetwarzania danych i informacji o ludziach w sposób automatyczny. W związku z tym realizacja zasad sprawiedliwości i słuszności w nowej rzeczywistości dominacji nieuchronnie związanej z technologicznym postępem wymaga zajęcia się podstawowymi społecznymi, ekonomicznymi i politycznymi źródłami tych problemów, a nie tylko oscylowania wokół nich[589].

Jak wskazuje Rediet Abebe, interwencja w poprawę funkcjonowania algorytmu, której celem jest ograniczenie ryzyka dyskryminacji przez ASD poprzez wyrównanie szans na przyjęcie na studia w różnych grupach demograficznych, może funkcjonować jako mniej ambitny substytut głębszej i trudniejszej pracy polegającej na poprawie nauczania w szkołach średnich w

[585] S. Fredman, *Intersectional Discrimination in EU Gender Equality and Non-Discrimination Law*, European Commission, Brussels 2016, s. 8.
[586] M. Mann, T. Matzner, *op. cit.*
[587] S. Barocas, A.D. Selbst, *op. cit.*, s. 676.
[588] M. Mann, T. Matzner, *op. cit.*
[589] R. Abebe *et al.*, *op. cit.*, s. 252.

182:24931

dzielnicach o niskich dochodach[590]. Podobnie koncentracja na interwencji w sposób działania ASD, polegająca np. na wyrównaniu szans zatrudnienia wśród kobiet i mężczyzn, może zmniejszyć wagę wrogiej kultury miejsca pracy lub innych barier utrudniających długoterminowy sukces pracownika, być może o podłożu nie tylko technologicznym (np. brak odpowiedniego systemu wsparcia i aktywizacji zawodowej kobiet po urodzeniu dziecka)[591]. W tym kontekście technologiczny wymiar procesu dyskryminacji może być jedynie nową ilustracją długotrwałych praktyk dominacji i segregacji społecznej. Dlatego też należy rozważyć wprowadzenie kategorii celu równości, który powinien być realizowany w ramach projektowanych rozwiązań proaktywnie wspierających zakaz dyskryminacji[592].

Eksploracja danych w systemach automatycznych może również wymagać ponownej oceny, dlaczego i w jakim zakresie należy realizować postulaty równościowe. Na poziomie Unii Europejskiej można wyróżnić cztery podstawowe sposoby rozumienia zasady równości:

> 1. równość jako nakaz usprawiedliwienia nierównego traktowania racjonalnymi przesłankami (równość jako racjonalność); 2. równość jako nakaz równej dystrybucji, szczególnie dóbr chronionych; 3. równość jako zakaz dyskryminacji ze względu na określone cechy szczególne; 4. równość jako pozytywny obowiązek promowania równości szans i równości materialnej[593].

W tym zakresie refleksja nad kierunkiem rozwoju prawa antydyskryminacyjnego z perspektywy ASD powinna rozwijać przede wszystkim ostatnie rozumienie równości, którego wymogiem jest kompensacja doznanych w przeszłości krzywd określonych grup, w przypadku których gwarancja równości formalnej prowadzi do nierównych rezultatów[594].

Chociaż dyskryminacja z pewnością trwa częściowo z powodu indywidualnych uprzedzeń decydentów, dużą część współczesnego krajobrazu nierówności można przypisać temu, co socjologowie nazywają instytucjonalną dyskryminacją[595]. Taka dyskryminacja wobec decentralizacji władzy przybiera jednak nowy wymiar. Z tej perspektywy niniejsza analiza ma na celu ukazanie, że problem dyskryminacji automatycznej to kwestia adaptacyjnego stosowania pojęcia dyskryminacji do rzeczywistości dominacji. Zaprezentowana praca ma się przyczynić nie tylko do poszerzenia dyskusji o opis zagrożeń stosowania ASD, ale też do

[590] *Ibidem*.
[591] *Ibidem*, s. 258.
[592] M. Mann, T. Matzner, *op. cit.*
[593] Ch. McCrudden, S. Prechal, *The Concepts of Equality and Non-Discrimination in Europe: A Practical Approach*, European Network of Legal Experts in the Field of Gender Equality 2009, https://ec.europa.eu/social/BlobServlet?docId=4553 (dostęp: 4.12.2021).
[594] A. Śledzińska-Simon, *op. cit.*, s. 53.
[595] S. Barocas, A.D. Selbst, *op. cit.*, s. 673.

183:11463

wskazania dróg umożliwiających szersze przeciwdziałanie procesom destrukcji społeczeństwa ze względu na pogłębiającą się na masową skalę, a przy tym niejawną marginalizację grup historycznie wykluczanych. W toku analizy procesów, które łączą się z wykluczeniem i marginalizacją, uwaga rządzących powinna koncentrować na tych, których społeczna i polityczna niewidzialność jest spotęgowana mimo istniejącej ochrony antydyskryminacyjnej[596]. Realizacja postulatów równościowych może przy tym wymagać aktywnego działania na rzecz obalenia ograniczających ją barier.

Podjęcie dalszych działań legislacyjnych w kierunku poszerzenia ochrony antydyskryminacyjnej zgodnie z założeniami teorii materialnej równości z pewnością doprowadzi do trudnej debaty społeczno-gospodarczej na temat relacji między prawem antydyskryminacyjnym a innymi prawami i wolnościami człowieka, w tym swobodami gospodarczymi. Z pewnością dyskusja ta zostanie poddana lobbingowi technologicznych gigantów, dla których ograniczenia w zakresie zbierania danych, sposobów ich przetwarzania i analizy w ASD, rynków zbytu automatycznych systemów decyzyjnych czy nowych form instytucjonalnej kontroli nad tym sektorem będzie równoznaczna z ograniczeniem potencjału informacyjnego i władzy zdobytej kosztem tradycyjnych form jej wykonywania.

O ile jednak antydyskryminacyjna ochrona gwarantowana przepisami prawa pozostanie dyskusją oscylującą wokół narzędzi ustanawianych w ramach formalnego postulatu równości, o tyle zastosowanie wobec dyskryminacji automatycznej realnej ochrony praw człowieka będzie niezwykle utrudnione i narażone na szybką dezaktualizację w obliczu zmieniającego się kształtu i kompetencji zaawansowanych algorytmów, w tym systemów sztucznej inteligencji.

[596] M. Mann, T. Matzner, *op. cit.*, s. 7.

184:26884

Zakończenie

Rozdział I niniejszej pracy dał odpowiedź na pytanie, czym jest dyskryminacja z perspektywy zarówno historyczno-politycznej, socjologicznej, jak i teoretycznoprawnej. Interdyscyplinarna analiza pojęciowa oraz rekonstrukcja podstawowych założeń zakazu dały mi możliwość zrozumienia przedmiotu regulacji prawa antydyskryminacyjnego. W toku tego badania doszłam do przekonania, że kontekstowe podejście jest kluczowe do oceny znaczenia automatycznego różnicowania dla dyskryminacji. Pojęcie dyskryminacji opiera się bowiem na zakazie różnicowania ze względu na cechy chronione, te zaś wykształcały się w sferze publicznej jako sprzeciw wobec dominacji przejawiającej się aktami fizycznymi.

W rozdziale II, na podstawie uprzednio ustalonego zakresu pojęcia dyskryminacji, przedstawiłam nowość procesu różnicowania automatycznego, w tym jego niejawność, nieprzejrzystość, masowość i złożoność, uniemożliwiające ludzki nadzór. Zaobserwowałam, że społeczno-polityczne tło stosunków dominacji uległo zmianie na tyle, że zakaz dyskryminacji przejawiający się penalizacją różnicowania ze względu na dotychczas wykształcone cechy chronione już nie wystarcza do zapewnienia ochrony praw i wolności człowieka. Jednocześnie doszłam do przeświadczenia, że wpływ algorytmów prowadzi aktualnie do wzmocnienia wykluczenia grup marginalizowanych i pokrzywdzonych, a skala niesprawiedliwości może wkrótce ulec dalszemu wzmocnieniu. Analiza zawarta w tym rozdziale doprowadziła mnie do wniosku, że rewolucja cyfryzacji i automatyzacji zmieniła obraz dominacji, przez co kontekst technologiczny wpływa na rozumienie pojęcia dyskryminacji.

W rozdziale III zbadałam, jak prawo funkcjonuje w warunkach transformacji i w zmienionych kontekstach normatywnych. Przedmiotem analizy doktrynalnej objęłam prawo antydyskryminacyjne, RODO oraz prawo do prywatności. Dostrzegłam, że wskazane regulacje są niedoskonałe i nieadekwatne do nowych jakościowo zjawisk technologicznych. Analiza stosowania prawa przez sądy i trybunały doprowadziła mnie do wniosku o istotnych ograniczeniach potencjalnej aktualizacji katalogu cech chronionych. W związku z tym uznaję, że różnicowanie automatyczne w swoim specyficznym kształcie, prowadzące do naruszenia praw i wolności człowieka na masową skalę, wymyka się odpowiedzi prawa. Możliwości normatywne obowiązującej definicji dyskryminacji i jej zakazu nie odpowiadają zatem na zmieniony kontekst polityczno-społeczny, przez co ryzyko postępującej marginalizacji i wykluczenia wzrasta.

185:74033

Rozdział IV zawiera z kolei odpowiedź na pytanie, czy współczesne prawodawstwo w zakresie projektów aktów prawnych regulujących sztuczną inteligencję prezentuje odpowiednio nowe jakościowo i rodzajowo zasady oceny ryzyka związanego z automatycznym różnicowaniem. Analiza projektu rozporządzenia dotyczącego tej kwestii doprowadziła mnie do wniosku, że brak ram teoretycznoprawnych w zakresie aktualnej treści definicji dyskryminacji w toku prawodawstwa regulującego cyfryzację może wpłynąć na upośledzenie regulacji, a nawet prowadzić do błędnego doboru metod legislacyjnych[597]. Refleksja nad zakresem pojęciowym dyskryminacji jest konieczna, gdyż wpisanie w nowe regulacje obecnego zakazu nie spowoduje objęcia penalizacją specyficznych form różnicowania algorytmicznego. Rozważania zawarte w tym rozdziale objęły również poszukiwanie odpowiedzi na pytanie, czy systemy automatyczne powinny być projektowane tak, aby móc interpretować przepisy samodzielnie, i jakie wyzwania wiążą się z technoregulacjami dla władzy ludzkiej. Wobec czasochłonności procesów legislacyjnych, a przy tym z uwagi na konieczność podjęcia szybkiej reakcji badanie doprowadziło mnie do wniosku o potrzebie teoretycznoprawnej aktualizacji pojęcia dyskryminacji w kontekście zmieniających się form dominacji w różnicowaniu automatycznym.

W rozdziale V zaś przedstawiłam kierunki możliwej ochrony przed różnicowaniem automatycznym, w tym poprzez rozwój prawa antydyskryminacyjnego i zarządzanie ryzykiem w automatycznych systemach decyzyjnych, opierając się na nowych regulacjach bazujących na „myśleniu ryzykiem". Zaprezentowałam również możliwość ukierunkowania definicji dyskryminacji bezpośredniej oraz pośredniej na zagrożenia związane ze specyfiką różnicowania automatycznego. Na podstawie zestawienia istotnych odmienności procesów różnicowania ludzkiego i automatycznego doszłam do przekonania o możliwości przedstawienia propozycji pojęcia dyskryminacji automatycznej jako niezależnej formy dyskryminacji. Badanie możliwych kierunków skłoniło mnie do przyjęcia, że ochrona jednostki przed dyskryminacją automatyczną powinna odbywać się w odniesieniu do postulatów równościowych nakierowanych na materialnoprawne wyrównywanie szans. Konieczna jest zatem dalsza refleksja teoretycznoprawna nad instrumentami prawa nie tylko z perspektywy rozwoju penalizacji dyskryminacji, ale też zapewnienia grupom historycznie marginalizowanym odpowiedniej reprezentacji politycznej, która umożliwi przerwanie kręgu wykluczenia.

[597] E.L. Sidorenko, P. von Arx, *op. cit*.

186:47742

Przeprowadzone w niniejszej pracy analizy potwierdzają, że automatyczne systemy decyzyjne, do których zaliczam algorytmy sztucznej inteligencji, decydując o prawach i obowiązkach człowieka, różnicują ludzi w sposób, który może prowadzić na masową skalę do wykluczenia grup społecznych historycznie marginalizowanych, a także wprowadzać nowe, nieoczywiste formy niesprawiedliwej kategoryzacji pozostałych obywateli. Algorytmy mogą przy tym różnicować zarówno podobnie do ludzi – powielając uprzedzenia historyczne zapisane w danych – jak i tworząc nowe formy niesprawiedliwości. Ta nowość może mieć przy tym dwa oblicza: 1. polegające na konstruowaniu grup poprzez identyfikację ludzi na podstawie cechy nieuznanej dotychczas za chronioną prawem antydyskryminacyjnym; 2. polegające na kategoryzowaniu w sposób niezrozumiały dla człowieka. W tym drugim wypadku nie chodzi o to, aby sama cecha była niepojmowalna, ale o to, że człowiek nie będzie w stanie odgadnąć przyczyny wyniku procesu decyzyjnego, a dokładnie – relewantnej dla procesu różnicowania cechy lub zespołu cech. Tak skategoryzowane przez algorytm grupy może łącznie charakteryzować heterogeniczny lub monolityczny dobór kryteriów różnicowania, społeczna doniosłość lub jej brak, a ponadto dokładność klasyfikacji lub jej brak[598]. Jednocześnie procesy różnicowania prowadzące do takiej wielopłaszczyznowej i skomplikowanej kategoryzacji mogą być niemoralne z uwagi na potęgowanie wykluczenia i upośledzenia grup historycznie doświadczających znacznej niesprawiedliwości.

O ile aktualnym celem zakazu dyskryminacji ma pozostać ograniczenie realnego problemu dominacji i systemowej opresji, o tyle dogłębna analiza mechanizmu powielania stereotypów i umacniania ich w danych może pozwolić na bliższe poznanie przyczyn formowania grup marginalizowanych. To zaś w dalszej perspektywie może nie tylko umożliwić efektywne przeciwdziałanie procesowi dyskryminacji, ale również doprowadzić do skonstruowania narzędzi, być może algorytmicznych, do wykrywania procesów prowadzących do wykluczenia i mityzacji takich zagrożeń na wcześniejszym etapie niż obecnie.

[598] S. Wachter, *op. cit.*, s. 13.

187:29584

Bibliografia

Abebe Rediet, Barocas Solon, Kleinberg Jon, Levy Karen, Raghavan Manish, Robinson David G., *Roles for Computing in Social Change*, [w:] *FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, [b.m.w.] 2020, s. 252–260, https://doi.org/10.1145/3351095.3372871.

Adamiak Barbara, Borkowski Janusz, *Kodeks postępowania administracyjnego. Komentarz*, C.H. Beck, Warszawa 2021.

Adams-Prassl Jeremias, Binns Reuben, Kelly-Lyth Aislinn, *Directly Discriminatory Algorithms*, „Modern Law Review" 86, 2022, nr 1, s. 144–175, https://doi.org/10.1111/1468-2230.12759.

Allhutter Doris, Cech Florian, Fischer Fabian, Grill Gabriel, Mager Astrid, *Algorithmic Profiling of Job Seekers in Austria: How Austerity Politics Are Made Effective*, „Frontiers in Big Data" 3, 2020, nr 5, https://doi.org/10.3389/fdata.2020.00005.

Alston Philip, *Report of the Special Rapporteur on Extreme Poverty and Human Rights*, United Nations General Assembly 2015, https://digitallibrary.un.org/record/798707.

Analiza wybranych aspektów projektu Aktu w sprawie sztucznej inteligencji, 2021, https://www.gov.pl/attachment/fa12287b-09c6-47d1-b0d6-7286c53d6917.

Barocas Solon, *Data Mining and the Discourse on Discrimination*, Center for Information Technology Policy, Princeton University, Princeton, NJ 2014, https://www.cs.yale.edu/homes/jf/Barocas-Taxonomy.pdf.

Barocas Solon, Selbst Andrew D., *Big Data's Disparate Impact*, Social Science Research Network, Rochester, NY 2016.

Barocas Solon, Hardt Moritz, Narayanan Arvind, *Fairness and Machine Learning*, https://fairmlbook.org.

Baroni Alessandro, Ruggieri Salvatore, *SCube: A Tool for Segregation Discovery*, [w:] *Advances in Database Technology – 22nd International Conference on Extending Database Technology*, EDBT, Lisbon 2019, s. 542–545, https://openproceedings.org/2019/conf/edbt/EDBT19_paper_212.pdf.

Barta Paweł, Dörre-Kolasa Dominika, Kawecki Maciej, Krzyżak Agnieszka, *Ogólne rozporządzenie o ochronie danych osobowych. Ustawa o ochronie danych osobowych. Wybrane przepisy sektorowe. Komentarz*, red. Paweł Litwiński, C.H. Beck, Warszawa 2021.

Baszkiewicz Jan, *1789: Spory o prawa człowieka i obywatela*, [w:] *Człowiek jako obywatel*, red. Maria Szyszkowska, Instytut Studiów Politycznych Polskiej Akademii Nauk, Warszawa 1995, s. 203–210.

Bauman Zygmunt, *Płynne czasy. Życie w epoce niepewności*, Wydawnictwo Sic!, Warszawa 2007.

———, *Retrotopia. Jak rządzi nami przeszłość?*, Wydawnictwo Naukowe PWN, Warszawa 2018.

———, *Życie na przemiał*, przeł. Tomasz Kunz, Wydawnictwo Literackie, Kraków 2005.

Bauman Zygmunt, Lyon David, *Płynna inwigilacja. Rozmowy*, przeł. Tomasz Kunz, Wydawnictwo Literackie, Kraków 2013.

188:79096

Bayamlıoğlu Emre, Leenes Ronald, *The "Rule of Law" Implications of Data-Driven Decision-Making: A Techno-Regulatory Perspective*, „Law, Innovation and Technology" 10, 2018, nr 2, s. 295–313.

Beauvoir Simone de, *Druga płeć*, t. 2, przeł. Maria Leśniewska, Wydawnictwo Literackie, Kraków 1972.

Beck Ulrich, *Społeczeństwo ryzyka. W drodze do innej nowoczesności*, przeł. Stanisław Cieśla, Wydawnictwo Naukowe Scholar, Warszawa 2002.

Berendt Bettina, *Algorithmic Discrimination*, [w:] *Encyclopedia of Law and Data Science*, 2020, https://people.cs.kuleuven.be/~bettina.berendt/Papers/Algorithmic%20Discrimination%20BB%202020%2010%2013.pdf.

———, *The AI Act Proposal: Towards the Next Transparency Fallacy? Why AI Regulation Should Be Based on Principles Based on How Algorithmic Discrimination Works*, [w:] *Künstliche Intelligenz – Wie gelingt eine vertrauenswürdige Verwendung in Deutschland und Europa?*, red. Frauke Rostalski, Mohr Siebeck, Tübingen 2022, s. 31–52.

Boles Janet K., Hoeveler Diane Long, *Historical Dictionary of Feminism*, Scarecrow Press, Lanham 2004.

Borgatta Edgar F., Montgomery Rhonda J.V., *Encyclopedia of Sociology*, Macmillan Reference USA, New York 2000.

Borgesius Frederik Zuiderveen, *Price Discrimination, Algorithmic Decision-Making, and European Non-Discrimination Law*, „European Business Law Review" 31, 2020, nr 3, s. 401–422, https://doi.org/10.54648/eulr2020017.

Bowker Geoffrey C., Star Susan Leigh, *Sorting Things Out. Classification and Its Consequences*, The MIT Press, Cambridge, MA 2000.

Box Goerge E. P., *Robustness in the strategy of scientific model building*, Robustness in Statistics, red. Robert Launer i Graham Wilkinson, Academic Press, 1979, s. 201–236.

Burek Wojciech, Klaus Witold, *Definiowanie dyskryminacji w prawie polskim w świetle prawa Unii Europejskiej oraz prawa międzynarodowego*, „Problemy Współczesnego Prawa Międzynarodowego, Europejskiego i Porównawczego" 11, 2013, s. 72–90, https://europeistyka.uj.edu.pl/documents/3458728/85399a2c-5e88-482a-8756-dc36bc715546.

Burrell Jenna, *How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms*, „Big Data & Society" 3, 2016, nr 1, s. 1–12.

Burt Andrew, *Is There a „Right to Explanation" for Machine Learning in the GDPR?*, IAPP 2017, https://iapp.org/news/a/is-there-a-right-to-explanation-for-machine-learning-in-the-gdpr/.

Cavanagh Matt, *Against Equality of Opportunity*, Oxford University Press, Oxford 2002.

Chander Sarah, Niklas Jędrzej, Veale Michael, Kaun Anne, Spurny Nina, Gollatz Benedikt *et al.*, *Q&A: How the EU's Flawed Artificial Intelligence Regulation Endangers the Social Safety Net*, Human Rights Watch 2021, https://www.hrw.org/news/2021/11/10/how-eus-flawed-artificial-intelligence-regulation-endangers-social-safety-net.

Chłopecki Aleksander, *Sztuczna inteligencja – szkice prawnicze i futurologiczne*, C.H. Beck, Warszawa 2021.

189:65882

Chomiczewski Witold, *Profilowanie – wyjaśnienie pojęcia*, [w:] *Polska i europejska reforma ochrony danych osobowych*, red. Dominik Lubasz, Edyta Bielak-Jomaa, Wolters Kluwer, Warszawa 2016, LEX/el.

Ciechomska Małgorzata, *Prawne aspekty profilowania oraz podejmowania zautomatyzowanych decyzji w ogólnym rozporządzeniu o ochronie danych osobowych*, „Europejski Przegląd Sądowy" 5, 2017, s. 37–42.

Cieśliński Aleksander, Szwarc Monika, Zawidzka-Łojek Anna, *Prawo rynku wewnętrznego. Funkcjonowanie Unii Europejskiej*, red. Dagmara Kornobis-Romanowska, „System Prawa Unii Europejskiej", t. 7, C.H. Beck, Warszawa 2019.

Corbett-Davies Sam, Goel Sharad, *The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning*, Cornell University, New York 2018, https://doi.org/10.48550/arXiv.1808.00023.

Crawford Kate, *Atlas of AI*, Yale University Press, New Haven-London 2021.

Crenshaw Kimberle, *Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics*, „The University of Chicago Legal Forum" 1, 1989, s. 139–176, https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=1052&context=uclf.

Criado Perez Caroline, *Niewidzialne kobiety*, przeł. Anna Sak, Karakter, Kraków 2020.

Cudowska Agata, *Tożsamość społeczna w „innej nowoczesności". Ulricha Becka koncepcja społeczeństwa ryzyka*, [w:] *Etniczność i obywatelskość w nowej Europie. Konteksty edukacji międzykulturowej*, red. Jerzy Nikitorowicz, Dorota Misiejuk, Mirosław Sobecki, Trans Humana Wydawnictwo Uniwersyteckie, Białystok 2007, s. 92–100.

Dastin Jeffrey, *Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women*, Reuters 2018, https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-airecruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.

DeFleur Melvin L., *"Stigma: Notes on the Management of Spoiled Identity" by Erving Goffman. Englewood Cliffs, New Jersey: Prentice-Hall, 1963. 147 pp. Cloth, $4.50; paper, $1.95*, „Social Forces" 43, 1964, nr 1, s. 127–128.

Derlatka Marta, Działocha Kazimierz, Jarosz-Żukowska Sylwia, Łukaszczuk Sylwia, Sarnecki Paweł, Sokolewicz Wojciech *et al.*, *Konstytucja Rzeczypospolitej Polskiej. Komentarz*, red. Leszek Garlicki, Marek Zubik, t. 2, Wydawnictwo Sejmowe, Warszawa 2016.

Domańska Monika, *Zakaz dyskryminacji ze względu na więcej niż jedno zabronione kryterium*, Wolters Kluwer, Warszawa 2019.

Dworkin Roland, *Biorąc prawa poważnie*, przeł. Jan Woleński, Wydawnictwo Naukowe PWN, Warszawa 1998.

Eckert Julia, *What Is the Context in "Law in Context"?*, [w:] *Concepts of Law. Comparative, Jurisprudential, and Social Science Perspectives*, red. Lukas Heckendorn Urscheler, Seán Patrick Donlan, Routledge, London 2014, s. 1–12, https://doi.org/10.4324/9781315573298.

Edwards Lilian, *The EU AI Act: A Summary of Its Significance and Scope*, Expert Explainer 2022, https://www.adalovelaceinstitute.org/wp-content/uploads/2022/04/Expert-explainer-The-EU-AI-Act-11-April-2022.pdf.

190:10330

An EU Artificial Intelligence Act for Fundamental Rights. A Civil Society Statement, 2021, https://edri.org/wp-content/uploads/2021/12/Political-statement-on-AI-Act.pdf.

Floridi Luciano, *Open Data, Data Protection, and Group Privacy*, „Philosophy & Technology" 27, 2014, nr 1, s. 1–3.

Foran Michael P., *Grounding Unlawful Discrimination*, „Legal Theory" 28, 2022, s. 3–34.

Fredman Sandra, *Discrimination Law*, Oxford University Press, Oxford 2002.

———, *Intersectional Discrimination in EU Gender Equality and Non-Discrimination Law*, European Commission, Brussels 2016.

Fredman Sandra, Szyszczak Erika M., *The Interaction of Race and Gender Discrimination*, [w:] *The Limits of Law*, red. Bob Hepple, Erika M. Szyszczak, Mansell, London-New York 1992, s. 214-226..

Freeman David, Ho Daniel E., Sharkey Catherine M., Cuéllar Mariano-Florentino, *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies*, Public Law Research Paper No. 20-54, NYU School of Law, New York 2020, http://dx.doi.org/10.2139/ssrn.3551505.

Fridman Lex, *MIT Deep Learning Basics: Introduction and Overview*, Massachusetts Institute of Technology, Cambridge, MA 2019, https://www.slideshare.net/noumfone/mit-deep-learning-basics-introduction-and-overview-by-lex-fridman.

Friedman Batya, Nissenbaum Helen, *Bias in Computer Systems*, „ACM Transactions on Information Systems" 14, 1996, nr 3, s. 330–347.

Gajane Pratik, Pechenizkiy Mykola, *On Formalizing Fairness in Prediction with Machine Learning*, Cornell University, New York 2018, https://arxiv.org/pdf/1710.03184.pdf.

Gandy Oscar H., *Coming to Terms with Chance: Engaging Rational Discrimination and Cumulative Disadvantage*, Taylor & Francis Ltd, Farnham 2009.

Gavison Ruth, *Privacy and the Limits of Law*, „The Yale Law Journal" 89, 1980, nr 3, s. 421–471, https://doi.org/10.2307/795891.

Gellert Raphaël, Vries Katia de, Hert Paul de, Gutwirth Serge, *A Comparative Analysis of Anti-Discrimination and Data Protection Legislations*, [w:] *Discrimination and Privacy in the Information Society. Studies in Applied Philosophy, Epistemology and Rational Ethics*, red. Bart Custers, Toon Calders, Bart Schermer, Tal Zarsky, t. 3, Springer, Berlin-Heidelberg, s. 61–89, https://doi.org/10.1007/978-3-642-30487-3_4.

Gerards Janneke, Xenidis Raphaële, *Algorithmic Discrimination in Europe: Challenges and Opportunities for Gender Equality and Non-Discrimination Law*, European Commission, Brussels 2020, https://op.europa.eu/en/publication-detail/-/publication/082f1dbc-821d-11eb-9ac9-01aa75ed71a1.

Gil Izabella, Głodowski Włodzimierz, Gołaczyński Jacek, Rodziewicz Piotr, Skibińska Magdalena, Skorupka Agnieszka, *Konwencja o ochronie praw człowieka i podstawowych wolności z dnia 4.11.1950*, red. Łukasz Błaszczak, t. 2, Legalis/el. 2021.

Gillespie Tarleton, *The Relevance of Algorithms*, [w:] *Media Technologies: Essays on Communication, Materiality, and Society Get Access Arrow*, red. Tarleton Gillespie, Pablo J. Boczkowski,

Kirsten A. Foot, MIT Press Scholarship Online, Cambridge, MA 2014, https://doi.org/10.7551/mitpress/9780262525374.003.0009.

Google LLC vs. Commission nationale de l'informatique et des libertés (CNIL), 2019, https://privacylibrary.ccgnlud.org/case/google-llc-vs-commission-nationale-de-linformatique-et-des-liberts-cnil.

Gouges Olympe de, *Déclaration Des Droits de La Femme et de La Citoyenne*, [b.m.w.] 1791.

Graumann Carl Friedrich, Wintermantel Margret, *Discriminatory Speech Acts: A Functional Approach*, [w:] *Stereotyping and Prejudice: Changing Conceptions*, red. Daniel Bar-Tal, Carl F. Graumann, Arie W. Kruglanski, Wolfgang Stroebe, Springer, New York 1989, s. 183–204.

Grzeszczak Robert, Szmigielski Artur, *Sądowe stosowanie Karty Praw Podstawowych UE w odniesieniu do państw członkowskich – refleksje na podstawie orzecznictwa Trybunału Sprawiedliwości i praktyki sądów krajowych*, „Europejski Przegląd Sądowy" 10, 2015, s. 11–18.

Handbook on European Nondiscrimination Law, European Union Agency for Fundamental Rights and Council of Europe, European Union and Council of Europe, Brussels 2018.

Hannam June, *Feminizm*, przeł. Agnieszka Kaflińska, Zysk i S-ka, Poznań 2010.

Harcourt Bernard E., *Against Prediction. Profiling, Policing, and Punishing in an Actuarial Age*, The University of Chicago Press, Chicago-London 2006.

Hassan Samer, De Filippi Primavera, *The Expansion of Algorithmic Governance: From Code Is Law to Law Is Code*, „Field Actions Science Reports" 17, 2017, s. 88–90.

Hatton Celia, *China 'Social Credit': Beijing Sets Up Huge System*, BBC 2015, https://www.bbc.com/news/world-asia-china-34592186.

Heinrichs Bert, *Discrimination in the Age of Artificial Intelligence*, „AI & Society" 37, 2022, s. 143–154, https://doi.org/10.1007/s00146-021-01192-2.

Hellman Deborah, *When Is Discrimination Wrong?*, Harvard University Press, Cambridge, MA-London 2011.

Hern Alex, *DeepMind Announces Ethics Group to Focus on Problems of AI*, „The Guardian" 4.10.2017, https://www.theguardian.com/technology/2017/oct/04/google-deepmind-ai-artificial-intelligence-ethics-group-problems.

Herek Gregory M., *Beyond "Homophobia": Thinking about Sexual Prejudice and Stigma in the Twenty-First Century*, „Sexuality Research & Social Policy" 1, 2004, s. 6–24.

———, *Science, Public Policy, and Legal Recognition of Same-Sex Relationships*, „American Psychologist" 62, 2007, nr 7, s. 713–715.

Hildebrandt Mireille, De Vries Katja, *Privacy, Due Process and the Computational Turn The Philosophy of Law Meets the Philosophy of Technology*, Routledge, London 2015.

Hill Kashmir, *Wrongfully Accused by an Algorithm*, „New York Times" 2020, https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html.

Hill Robin K., *What an Algorithm Is*, „Philosophy & Technology" 29, 2016, nr 1, s. 35–59.

192:87125

Hoven van Genderen Robert van den, *Do We Need New Legal Personhood in the Age of Robots and AI?*, [w:] *Robotics, AI and the Future of Law*, red. Marcelo Corrales, Mark Fenwick, Nikolaus Forgó, Springer, Singapore 2018, s. 15–50.

Hryniewicz Janusz T., *Społeczeństwo ryzyka. Teoria, model, analiza krytyczna*, „Przegląd Socjologiczny" 2, 2014, s. 9–33.

Huhns Michael N., *Being and Acting Rational [Agent Design]*, „IEEE Internet Computing" 7, 2003, nr 2, s. 91–93.

Humphreys Paul, *The Philosophical Novelty of Computer Simulation Methods*, „Synthese" 169, 2009, nr 3, s. 615–626.

Ingold David, Soper Spencer, *Amazon Doesn't Consider the Race of Its Customers. Should It?*, Bloomberg 2016, https://www.bloomberg.com/graphics/2016-amazon-same-day/?cmpid=google.

Ingram David, *World Crisis and Underdevelopment A Critical Theory of Poverty, Agency, and Coercion*, Cambridge University Press, Cambridge, UK 2018.

Ishwara Bhat P., *Idea and Methods of Legal Research*, Oxford University Press, New Delhi 2019.

Jacobs Adam, *The Pathologies of Big Data*, „Queue" 7, 2009, nr 6, s. 10–19, https://doi.org/10.1145/1563821.1563874.

Johnson Deborah G., Nissenbaum Helen (red.), *Computers, Ethics & Social Values*, Prentice Hall, Upper Saddle River, NJ 1995.

Kański Lesław, *Prawo do prywatności*, [w:] *Prawa człowieka. Model prawny*, red. Roman Wieruszewski, Zakład Narodowy im. Ossolińskich–Wydawnictwo Polskiej Akademii Nauk, Warszawa 1991. s. 326-329.

Kearns Michael, Neel Seth, Roth Aaron, Wu Zhiwei Steven, *Preventing Fairness Gerrymandering: Auditing and Learning for Subgroup Fairness. Proceedings of the 35th International Conference on Machine Learning*, „Proceedings of Machine Learning Research" 80, 2018, s. 2564–2572, https://proceedings.mlr.press/v80/kearns18a.html.

Khaitan Tarunabh, *Prelude to a Theory of Discrimination Law*, [w:] *Philosophical Foundations of Discrimination Law*, red. Deborah Hellman, Sophia Moreau, Oxford University Press, Oxford 2013, s. 138–162.

Kibil Michał, *Profilowanie i automatyczne podejmowanie decyzji*, LEX/el. 2018.

Kim Pauline T., *Data-Driven Discrimination at Work*, „William & Mary Law Review" 58, 2017, nr 3, https://scholarship.law.wm.edu/wmlr/vol58/iss3/4.

Klaus Witold, *Prawo jako instrument mający zapewnić równość w społeczeństwie*, [w:] *Różni, ale równi. Prawo a równe traktowanie cudzoziemców w Polsce*, red. Witold Klaus, Stowarzyszenie Interwencji Prawnej, Warszawa 2013, s. 5–19.

Kopff Andrzej, *Koncepcja praw do intymności i do prywatności życia osobistego*, „Studia Cywilistyczne" 20, 1972, s. 32–33.

Koranyi Karol, *Powszechna historia państwa i prawa*, t. 4, Państwowe Wydawnictwo Naukowe, Warszawa 1967.

193:37687

Krasuski Andrzej, *Chmura obliczeniowa. Prawne aspekty zastosowania*, Wolters Kluwer, Warszawa 2017.

———, *Status prawny sztucznego agenta. Podstawy prawne zastosowania sztucznej inteligencji*, C.H. Beck, Warszawa 2020.

Kułak Maciej, Śmiszek Krzysztof, *Kilka uwag o standardzie równości w Europejskiej Konwencji Praw Człowieka*, „Europejski Przegląd Sądowy" 2, 2017, s. 40–45.

Kwiatkowska Ewa M., *Rozwój Internetu rzeczy – szanse i zagrożenia*, „Internetowy Kwartalnik Antymonopolowy i Regulacyjny" 8, 2014, nr 3, s. 60–70.

Leese Matthias, *The New Profiling: Algorithms, Black Boxes, and the Failure of Anti-Discriminatory Safeguards in the European Union*, „Security Dialogue" 45, 2014, s. 494–511.

Lerman Jonas, *Big Data and Its Exclusions*, „Stanford Law Review" 66, 2013, https://www.stanfordlawreview.org/online/privacy-and-big-data-big-data-and-its-exclusions/.

Leslie Sarah-Jane, *Generics Oversimplified*, „NOUS" 49, 2015, s. 28–54.

Lessig Lawrence, *Code: And Other Laws of Cyberspace, Version 2.0*, Basic Books, New York 2006.

Lippert-Rasmussen Kasper, *Born Free and Equal? A Philosophical Inquiry into the Nature of Discrimination*, Oxford University Press, Oxford 2013.

———, *The Badness of Discrimination*, „Ethical Theory and Moral Practice" 9, 2006, nr 2, s. 167–185.

Liżewski Bartosz, *System prawa a porządek prawny. Między konstrukcją normatywną a prawem w działaniu*, C.H. Beck, Warszawa 2021.

Lubasz Dominik, Big Brother Is Watching You. *Profilowanie i zautomatyzowane podejmowanie decyzji w kontekście zasad legalności i przejrzystości*, [w:] *Rok RODO 2019*, red. Wojciech Wiewiórowski, Hanna Wolska, C.H. Beck, Warszawa 2019, s. 115–134.

———, *Zasady legalności, przejrzystości i minimalizacji danych w ogólnym rozporządzeniu o ochronie danych osobowych w kontekście sztucznej inteligencji*, [w:] *Prawo sztucznej inteligencji*, red. Luigi Lai, Marek Świerczyński, C.H. Beck, Warszawa 2020, s. 173–186.

Lubasz Dominik, Namysłowska Monika, *Zasady dotyczące przetwarzania danych osobowych a sztuczna inteligencja w kontekście europejskim*, [w:] *Sztuczna inteligencja*, blockchain, *cyberbezpieczeństwo oraz dane osobowe. Zagadnienia wybrane*, red. Kinga Flaga-Gieruszyńska, Jacek Gołaczyński, Dariusz Szostek, C.H. Beck, Warszawa 2019, Legalis/el.

Makkonen Timo, *Hauptursachen, Formen und Folgen von Diskriminierung*, [w:] *Handbuch zur Rechtlichen Bekämpfung von Diskriminierung*, Internationale Organisation Für Migration (IOM), Regionalbüro Für Die Baltischen Und Nordischen Staaten, Helsinki 2003, s. 8–31.

———, *Multiple, Compound and Intersectional Discrimination: Bringing the Experiences of the Most Marginalized to the Fore*, Åbo Akademi University, Institute for Human Rights, Turku 2002, https://www.abo.fi/wp-content/uploads/2018/03/2002-Makkonen-Multiple-compound-and-intersectional-discrimination.pdf.

Maliszewska-Nienartowicz Justyna, *Karta Praw Podstawowych Unii Europejskiej. Komentarz*, red. Andrzej Wróbel, C.H. Beck, Warszawa 2020.

194:60510

———, *Prawo antydyskryminacyjne*, red. Justyna Maliszewska-Nienartowicz, „System Prawa Unii Europejskiej", t. 6, C.H. Beck, Warszawa 2020.

Mann Monique, Matzner Tobias, *Challenging Algorithmic Profiling: The Limits of Data Protection and Anti-Discrimination in Responding to Emergent Discrimination*, „Big Data & Society" 6, 2019, nr 2, s. 1–11, https://doi.org/10.1177/20539517198958.

Maroń Grzegorz, *Wstęp do prawoznawstwa*, Wydawnictwo Uniwersytetu Rzeszowskiego, Rzeszów 2011.

Mattioli Michael, *Disclosing Big Data*, „Minnesota Law Review" 99, 2014, s. 535–583.

McCrudden Christopher, Prechal Sacha, *The Concepts of Equality and Non-Discrimination in Europe: A Practical Approach*, European Network of Legal Experts in the Field of Gender Equality 2009, https://ec.europa.eu/social/BlobServlet?docId=4553.

Mednis Arwid, *Istota profilowania osób i jego wykorzystanie w sektorze publicznym*, „ABI-Expert" 1, 2018, http://www.abi-expert.pl/wydania/styczen-marzec-2018/art,1986,istota-profilowania-osob-i-jego-wykorzystanie-w-sektorze-publicznym.html.

———, *Prawo do prywatności a interes publiczny*, Wolters Kluwer–Kantor Wydawniczy Zakamycze, Warszawa-Kraków 2006.

Mercat-Bruns Marie, *Multiple Discrimination and Intersectionality: Issues of Equality and Liberty*, „International Social Science Journal" 67, 2018, nr 223–224, s. 43–54.

Merrills J.G., Robertson A.H., *Human Rights in Europe*, Juris Publishing, Inc., New York-Manchester, UK 2001.

Michalak Arkadiusz, *Odpowiedzialność cywilnoprawna w obrocie oprogramowaniem komputerowym w erze sztucznej inteligencji*, C.H. Beck, Warszawa 2021.

Michałowska Kinga, *Różnicowanie a dyskryminacja – status prawny osób szczepionych i nieszczepionych na COVID-19*, „Przegląd Sądowy" 4, 2021, s. 38–56.

Miller Alex P., *Decision Making and Problem Solving Want Less-Biased Decisions? Use Algorithms*, „Harvard Business Review" 2018, https://hbr.org/2018/07/want-less-biased-decisions-use-algorithms.

Mitchell Tom M., *Machine Learning*, McGraw-Hill, New York 1997.

Mittelstadt Brent, Allo Patrick, Taddeo Mariarosaria, Wachter Sandra, Floridi Luciano, *The Ethics of Algorithms: Mapping the Debate*, „Big Data & Society. In press" 2, 2016, s. 1–21, https://doi.org/10.1177/2053951716679679.

Mittelstadt Brent Daniel, *From Individual to Group Privacy in Big Data Analytics*, „Philosophy & Technology" 30, 2017, s. 475–495.

Moka-Mubelo S.J. Willy, *Towards a Contextual Understanding of Human Rights*, „Ethics & Global Politics" 12, 2019, nr 4, s. 40–52.

Mokhtarian Edmund, *The Bot Legal Code: Developing a Legally Compliant Artificial Intelligence*, „Vanderbilt Journal of Entertainment & Technology Law" 21, 2018, s. 145–207.

Moreau Sophia, *Discrimination Law and the Freedom to Live a Good Life*, „Law and Philosophy" 35, 2016, s. 511–527.

195:43681

Morondo Taramundi Dolores, *Discrimination by Machine-Based Decisions: Inputs and Limits of Anti-Discrimination Law*, [w:] *Law and Artificial Intelligence: Regulating AI and Applying AI in Legal Practice*, red. Bart Custers, Eduard Fosch-Villaronga, Springer, The Hague 2022, s. 73–85.

Muñoz Cecilia, Smith Megan, Patil DJ, *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights*, Executive Office of the President 2016, https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf.

Niezależna grupa ekspertów wysokiego szczebla ds. sztucznej inteligencji powołana przez Komisję Europejską w czerwcu 2018 r., *Wytyczne w zakresie etyki dotyczące godnej zaufania sztucznej inteligencji*, Komisja Europejska, Bruksela 2019, https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/JURI/DV/2019/11-06/Ethics-guidelines-AI_PL.pdf.

Niklas Jędrzej, *Problem dyskryminacji automatycznej – uwagi na tle ogólnego rozporządzenia o ochronie danych osobowych*, „Europejski Przegląd Sądowy" 7, 2019, s. 4–9.

———, *Profilowanie w kontekście ochrony danych osobowych i zakazu dyskryminacji*, Polskie Towarzystwo Prawa Antydyskryminacyjnego [b.r.w.], http://ptpa.org.pl/site/assets/files/publikacje/opinie/Opinia_profilowanie_w_kontekscie_ochrony_danych_osobowych_i_zakazu_dyskryminacji.pdf.

Niklas Jędrzej, Sztandar-Sztanderska Karolina, Szymielewicz Katarzyna, *Profiling the Unemployed in Poland: Social and Political Implications of Algorithmic Decision Making*, Fundacja Panoptykon, Warszawa 2015, https://panoptykon.org/sites/default/files/leadimage-biblioteka/panoptykon_profiling_report_final.pdf.

Nowakowski Michał, *FINTECH – technologia, finanse, regulacje. Praktyczny przewodnik dla sektora innowacji finansowych*, Wolters Kluwer, Warszawa 2020.

Nowicki Marek Antoni, *Komentarz do Konwencji o ochronie praw człowieka i podstawowych wolności*, [w:] *Wokół Konwencji Europejskiej. Komentarz do Europejskiej Konwencji Praw Człowieka*, Wolters Kluwer, Warszawa 2021, https://sip.lex.pl/komentarze-i-publikacje/komentarze/komentarz-do-konwencji-o-ochronie-praw-czlowieka-i-podstawowych-587259731.

Ntoutsi Eirini, Fafalios Pavlos, Gadiraju Ujwal, Iosifidis Vasileios, Nejdl Wolfgang, Vidal Maria-Esther, Ruggieri Salvatore, Turini Franco, Papadopoulos Symeon, Krasanakis Emmanouil, Kompatsiaris Ioannis, Kinder-Kurlanda Katharina, Wagner Claudia, Karimi Fariba, Fernandez Miriam, Alani Harith, Berendt Bettina, Kruegel Tina, Heinze Christian, Broelemann Klaus, Kasneci Gjergji, Tiropanis Thanassis, Staab Steffen, *Bias in Data-Driven Artificial Intelligence Systems – An Introductory Survey*, „WIREs Data Mining and Knowledge Discovery" 10, 2020, nr 3, e1356, s. 1–14.

O'Neil Cathy, *Broń matematycznej zagłady: jak algorytmy zwiększają nierówności i zagrażają demokracji*, przeł. Marcin Z. Zieliński, PWN, Warszawa 2020.

———, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Crown Publishers, New York 2017.

Orwat Carsten, *Risks of Discrimination through the Use of Algorithms*, Federal Anti-Discrimination Agency 2019,

196:11167

https://www.antidiskriminierungsstelle.de/EN/homepage/_documents/download_diskr_risiken_verwendung_von_algorithmen.pdf?__blob=publicationFile&v=1.

Phillips P. Jonathon, Hahn Carina A., Fontana Peter C., Broniatowski David A., Przybocki Mark A., *Four Principles of Explainable Artificial Intelligence*, National Institute of Standards and Technology, U.S. Department of Commerce 2020, https://nvlpubs.nist.gov/nistpubs/ir/2021/NIST.IR.8312.pdf.

Pichlak Maciej, Gaczoł Klaudia, *Refleksyjna regulacja ryzyka w ochronie danych osobowych. Podstawy teoretyczne i praktyka działalności prezesa UODO*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2022.

Pincus Fred L., *Discrimination Comes in Many Forms: Individual, Institutional, and Structural*, „The American Behavioral Scientist" 40, 1996, nr 2, s. 186–194.

Płonowska-Ziarek Ewa, *Rządy ludzkie czy algorytmiczne? O automatyzacji władzy sądzenia*, wykład mistrzowski, Uniwersytet Jagielloński, Kraków 2021.

Pollicino Oreste, De Gregorio Giovanni, *Constitutional Law in the Algorithmic Society*, [w:] *Constitutional Challenges in the Algorithmic Society*, red. Hans W. Micklitz, Oreste Pollicino, Amnon Reichman, Andrea Simoncini, Giovanni Sartor, Giovanni De Gregorio, Cambridge University Press, Cambridge, MA 2021, s. 3–24, https://doi.org/10.1017/9781108914857.002.

Polonetsky Jules, Tene Omer, *Taming the Golem: Challenges of Ethical Algorithmic Decision Making*, „North Carolina Journal of Law & Technology" 19, 2018, nr 1, s. 125–173.

Popper Jens *et al.*, *Artificial Intelligence across Industries*, IEC White Paper 2018, https://www.researchgate.net/publication/329191549_Artificial_intelligence_across_industries_-_IEC_Whitepaper.

Porzeżyński Marek, *Problematyka własności danych generowanych w ramach Internetu rzeczy*, „Państwo i Prawo" 2, 2021, s. 123–137.

Presidency Conclusions, The Charter of Fundamental Rights in the Context of Artificial Intelligence and Digital Change, https://polit-x.de/en/documents/4580568/.

Purtova Nadezhda, *The Law of Everything: Broad Concept of Personal Data and Future of EU Data Protection Law*, „Law, Innovation and Technology" 10, 2018, nr 1, s. 40–81, https://doi.org/10.1080/17579961.2018.1452176.

Q&A: How the EU's Flawed Artificial Intelligence Regulation Endangers the Social Safety Net, https://www.hrw.org/news/2021/11/10/how-eus-flawed-artificial-intelligence-regulation-endangers-social-safety-net.

Reisigl Martin, *Dyskryminacja w dyskursach*, „Tekst i Dyskurs – Text und Diskurs" 3, 2010, s. 27–61.

Robin-Olivier Sophie, *Protection of Fundamental Rights under Internal Market Rules and in the Area of Freedom, Security and Justice: Unity or Dualism?*, University of Paris Quest-Naterre La Défense – Working Paper, Paris 2009.

Rojszczak Marcin, *Ochrona prywatności w cyberprzestrzeni z uwzględnieniem zagrożeń wynikających z nowych technik przetwarzania informacji*, Wolters Kluwer, Warszawa 2019.

———, *Prawne aspekty systemów sztucznej inteligencji – zarys problemu*, [w:] *Sztuczna inteligencja, blockchain, cyberbezpieczeństwo oraz dane osobowe. Zagadnienia wybrane*, red. Kinga Flaga-Gieruszyńska, Jacek Gołaczyński, Dariusz Szostek, C.H. Beck, Warszawa 2019, Legalis/el.

197:85986

Ryszka Joanna, *Prawa podstawowe jako uzasadnione ograniczenie swobód rynku wewnętrznego Unii Europejskiej*, „Przegląd Prawa i Administracji" 96, 2014, s. 59–78.

Sadurski Wojciech, *Równość wobec prawa*, „Państwo i Prawo" 8–9, 1978, s. 51–64.

Sakowska-Baryła Marlena (red.), *Ogólne rozporządzenie o ochronie danych osobowych*, C.H. Beck, Warszawa 2018.

Sandvig Christian, Hamilton Kevin, Karahalios Karrie, Langbort Cédric, *Auditing Algorithms : Research Methods for Detecting Discrimination on Internet Platforms*, [w:] *Data and Discrimination: Converting Critical Concerns into Productive: A Preconference at the 64th Annual Meeting of the International Communication Association*, Seattle 2014, https://www.kevinhamilton.org/share/papers/Auditing%20Algorithms%20--%20Sandvig%20--%20ICA%202014%20Data%20and%20Discrimination%20Preconference.pdf.

Schäfer Marie, Haun Daniel B.M., Tomasello Michael, *Fair Is Not Fair Everywhere*, „Psychological Science" 26, 2015, nr 8, s. 1252–1260.

Schauer Frederick, *Profiles, Probabilities, and Stereotypes*, The Belknap Press of Harvard University Press, London 2003.

Schiek Dagmar, Lawson Anna, *European Union Non-Discrimination Law and Intersectionality, Investigating the Triangle of Racial, Gender and Disability Discrimination*, Routledge, London 2011.

Schuett Jonas, *Risk Management in the Artificial Intelligence Act*, „European Journal of Risk Regulation" 2023, s. 1–19, https://doi.org/10.1017/err.2023.1.

Shin Patrick S., *Is There a Unitary Concept of Discrimination?*, [w:] *Philosophical Foundations of Discrimination Law*, red. Deborah Hellman, Sophia Moreau, Oxford University Press, Oxford 2013, s. 163–181.

Shirrell Stuart, *The Privacy Pro's Guide to Explainability in Machine Learning*, IAPP, https://iapp.org/news/a/the-privacy-pros-guide-to-explainability-in-machine-learning/.

Sidorenko Elina L., Arx Pierre von, *Transformation of Law in the Context of Digitalization: Defining the Correct Priorities*, „Digital Law Journal" 1, 2020, nr 1, s. 24–38, https://doi.org/10.38044/DLJ-2020-1-1-24-38.

Skalfist Peter, Mikelsten Daniel, Teigens Vasil, *Sztuczna inteligencja: czwarta rewolucja przemysłowa*, przeł. C.S.B. Equipment, Cambridge Stanford Books, [b.m.w.] 2020.

Stangor Charles, Macrae C. Neil, Hewstone Miles, *Stereotypy i uprzedzenia*, przeł. Małgorzata Majchrzak [*et al.*], Gdańskie Wydawnictwo Psychologiczne, Gdańsk 1999.

Sunstein Cass R., *Sprzeciw w życiu społeczeństw*, przeł. Jerzy S. Kugler, Wydawnictwo Sejmowe, Warszawa 2006.

Syska Katarzyna, *Ocena odpowiedniości przepisów RODO do zapewnienia przejrzystości działania systemów AI – wybrane zagadnienia*, [w:] *Ocena i przegląd RODO po dwóch latach obowiązywania. Aktualne problemy prawnej ochrony danych osobowych 2020*, red. Grzegorz Sibiga, dodatek specjalny do „Monitora Prawniczego" 23, 2020, s. 77-85..

Szeliga Marcin, Data Science *i uczenie maszynowe*, Wydawnictwo Naukowe PWN, Warszawa 2017.

198:77615

Szostek Dariusz, *Is the Traditional Method of Regulation (the Legislative Act) Sufficient to Regulate Artificial Intelligence, or Should It Also Be Regulated by an Algorithmic Code?*, „Bialystok Legal Studies" 26, 2021, nr 3, s. 43–60.

———, *LegalTech. Czyli jak bezpiecznie korzystać z narzędzi informatycznych w organizacji, w tym w kancelarii oraz dziale prawnym*, C.H. Beck, Warszawa 2021.

———, *Sztuczna inteligencja a kody. Czy rozwiązaniem dla uregulowania sztucznej inteligencji jest* Smart Contract *i* Blockchain*?*, [w:] *Prawo sztucznej inteligencji*, red. Luigi Lai, Marek Świerczyński, C.H. Beck, Warszawa 2020, Legalis/el.

Śledzińska-Simon Anna, *Zasada równości i zasada niedyskryminacji w prawie Unii Europejskiej*, „Studia BAS" 2, 2011, nr 26, s. 41–86.

Taekema Sanne, Burg Wibren van der, *Legal Philosophy as an Enrichment of Doctrinal Research Part I: Introducing Three Philosophical Methods*, „Law and Method" 2020, s. 1–19, http://www.bjutijdschriften.nl/tijdschrift/lawandmethod/2020/01/lawandmethod-D-19-00006.

Thomsen Frej Klem, *But Some Groups Are More Equal than Others: A Critical Review of the Group-Criterion in the Concept of Discrimination*, „Social Theory and Practice" 39, 2013, nr 1, s. 120–146.

Thomsen Frej Klem, Khaitan Tarunabh, Schauer Frederick, *The Routledge Handbook of the Ethics of Discrimination*, red. Kasper Lippert-Rasmussen, Routledge, New York-London 2017.

Tobler Christa, *Limits and Potential of the Concept of Indirect Discrimination*, European Commission, Directorate-General for Employment, Social Affairs and Inclusion 2008, https://data.europa.eu/doi/10.2767/56607.

Toh Amos, *Dutch Ruling a Victory for Rights of the Poor*, Human Rights Watch 2020, https://www.hrw.org/news/2020/02/06/dutch-ruling-victory-rights-poor.

Trzciński Krzysztof, *Obywatelstwo w Europie. Idea i jej wyraz formalny w perspektywie historycznej*, „Studia Europejskie" 2, 2002, s. 46–55.

———, *Początki nowożytnego obywatelstwa w Europie – obywatel państwa i katalog jego praw w dokumentach Rewolucji Francuskiej*, „Studia Europejskie" 2, 2005, s. 67–94.

Turilli Matteo, *Ethical Protocols Design*, „Ethics and Information Technology" 9, 2007, nr 1, s. 49–62, https://doi.org/10.1007/s10676-006-9128-9.

Turner Jacob, *Robot Rules: Regulating Artificial Intelligence*, Palgrave Macmillan, Cham 2019, https://doi.org/10.1007/978-3-319-96235-1.

Turska Anna, *Prawo i wykluczenie. Studium empiryczne*, C.H. Beck, Warszawa 2018.

Vandenhole Wouter, *Non-Discrimination and Equality in the View of the UN Human Rights Treaty Bodies*, Intersentia, Oxford 2005.

Veale Michael, Zuiderveen Borgesius Frederik, *Demystifying the Draft EU Artificial Intelligence Act. Analysing the Good, the Bad, and the Unclear Elements of the Proposed Approach*, „Computer Law Review International" 4, 2021, s. 97–112.

Wachter Sandra, *The Theory of Artificial Immutability: Protecting Algorithmic Groups under Anti-Discrimination Law*, „Tulane Law Review" 20, 2022, nr 97, s. 1–50, http://dx.doi.org/10.2139/ssrn.4099100.

199:70034

Warren Samuel D., Brandeis Louis D., *The Right to Privacy*, „Harvard Law Review" 4, 1890, nr 5, s. 193–220.

Werbach Kevin, *The Blockchain and the New Architecture of Trust*, The MIT Press, Cambridge, MA 2018.

Wiewiórowski Wojciech, *Profilowanie osób na podstawie ogólnodostępnych danych*, [w:] *Prywatność a ekonomia. Ochrona danych osobowych w obrocie gospodarczym*, red. Arwid Mednis, Wydawnictwa Uniwersytetu Warszawskiego, Warszawa 2013, s. 25–40.

Winiarska Aleksandra, Klaus Witold, *Dyskryminacja i nierówne traktowanie jako zjawisko społeczno-kulturowe*, „Studia BAS" 2, 2011, nr 26, s. 9–39.

Wiórek Piotr Marcin, *Pojęcie sztucznej inteligencji*, [w:] *Sto lat polskiego prawa handlowego. Księga jubileuszowa dedykowana Profesorowi Andrzejowi Kidybie*, red. Małgorzata Dumkiewicz, Jerzy Szczotka, Katarzyna Kopaczyńska-Pieczniak, t. 1, Wolters Kluwer, Warszawa 2020, LEX/el.

Wollstonecraft Mary, *A Vindication of the Rights of Woman*, Peter Edes for Thomas and Andrews, Faust's Statue, Boston 1792.

Wróbel Andrzej, *Dyskryminacja bezpośrednia ze względu na więź?*, „Europejski Przegląd Sądowy" 9, 2008, LEX/el..

———, *Konwencja o ochronie praw człowieka i podstawowych wolności*, t. 1. *Komentarz do artykułów 1–18*, red. Leszek Garlicki, C.H. Beck, Warszawa 2010.

Wyrzykowski Mirosław, Ziółkowski Michał, *Konstytucyjny status proceduralny jednostki jako adresata działań organów administracji*, red. Roman Hauser, Andrzej Wróbel, Zygmunt Niewiadomski, „System Prawa Administracyjnego", t. 2, C.H. Beck, Warszawa 2012.

Xenidis Raphaële, Senden Linda, *EU Non-Discrimination Law in the Era of Artificial Intelligence: Mapping the Challenges of Algorithmic Discrimination*, [w:] *General Principles of EU Law and the EU Digital Order*, red. Ulf Bernitz, Xavier Groussot, Jaan Paju, Sybe A. De Vries, Wolters Kluwer, Alphen aan den Rijn 2020, s. 151–181.

Young Iris Marion, *Justice and the Politics of Difference*, Princeton University Press, Princeton, NJ 2022.

Zalewski Tomasz, *Definicja sztucznej inteligencji*, [w:] *Prawo sztucznej inteligencji*, red. Luigi Lai, Marek Świerczyński, C.H. Beck, Warszawa 2020, Legalis/el.

Zinn Howard, *Ludowa historia Stanów Zjednoczonych. Od roku 1492 do dziś*, przeł. Andrzej Wojtasik, Wydawnictwo Krytyki Politycznej, Warszawa 2016.

Zuboff Shoshana, *Wiek kapitalizmu inwigilacji*, przeł. Alicja Unterschuetz, Zysk i S-ka, Poznań 2020.

Zuiderveen Borgesius Frederik, *Price Discrimination, Algorithmic Decision-Making, and European Non-Discrimination Law*, „European Business Law Review" 2020, s. 1–29.

200:38241