

Marta Otto  <https://orcid.org/0000-0002-8970-9643>

Uniwersytet Łódzki

DYSKRYMINACJA ALGORYTMICZNA W ZATRUDNIENIU. ZARYS PROBLEMU

Abstract

Algorithmic discrimination in employment. An overview

Debates about the future of work in light of developments in artificial intelligence are held predominantly in the context of job losses and technological unemployment. Far less attention is paid to the challenges posed by the increasingly widespread phenomenon of algorithmisation of management functions in the modern world of work. Meanwhile, the shift to algorithmic management represents a significant qualitative change that, in addition to promising broadly understood modernisation and optimisation of decision-making processes, carries specific repercussions in the context of human rights protection, including in particular the prohibition of discrimination in employment. The article attempts to assess the adequacy of the EU anti-discrimination instrumentarium to the specifics of algorithmic discrimination mechanisms, and aims to encourage an in-depth scientific discussion on the need to develop effective regulatory responses at the national level.

Słowa kluczowe: zarządzanie algorytmiczne, dyskryminacja algorytmiczna, unijne prawo antydyskryminacyjne

Keywords: algorithmic management, algorithmic discrimination in employment, EU non-discrimination law

ASJC: 3308, **JEL:** K31

Uwagi wprowadzające

Oficjalnie zainaugurowany w 1956 roku przez grupę informatyków w ramach szkoły letniej w Dartmouth projekt sztuczna inteligencja (*artificial intelligence*, AI) zdaje się przeżywać obecnie swoje *momentum*, również w sferze zatrudnienia. Spektakularny postęp, jaki dokonał się w tej dziedzinie na przestrzeni ostatnich lat, m.in. dzięki wykładniczemu rozwojowi nowych technologii umożliwiającym bezprecedensowy dostęp do różnego rodzaju danych, doprowadził do upowszechnienia się algorytmów w różnych obszarach życia społecznego. Inteligentne systemy są wdrażane w wielu sektorach – od zdrowia,

finansów czy sądownictwa po ubezpieczenia i reklamę. Współczesny świat pracy odchodzi przy tym od spolaryzowanego scenariusza, w którym wysoce manualne zawody z jednej strony i zawody oparte na umiejętnościach z drugiej wydawały się niezagrożone automatyzacją. Wręcz przeciwnie, to, co zaczęło się od algorytmizacji konkretnych zadań zarządczych w tzw. *gig economy* (Ajunwa, Greene 2018), rozprzestrzenia się dziś na cały świat pracy – jesteśmy świadkami gwałtownego wzrostu zarówno pod względem zakresu miejsc pracy, gdzie wdrażane są inteligentne technologie, jak i funkcji zarządczych, które można zautomatyzować (Wood 2021, s. 6–7).

Rozwój zarządzania algorytmicznego (znanego również jako „People Analytics” lub „Big Data HR”) – tj. coraz bardziej wszechobecnej zależności od technologii monitorowania i wyrafinowanych algorytmów do pomiaru, kontroli i oceny pracowników (Otto 2019) – pozwala na sprawowanie dotychczas niemożliwej do osiągnięcia wielowymiarowej kontroli nad pracownikami. Zarządzanie algorytmiczne wydaje się przy tym swoistego rodzaju powrotem do teorii „naukowego zarządzania” Fredericka Winsłowa Taylora (1972), ale bez znaczących kosztów transakcyjnych, które wcześniej stanowiły naturalne ograniczenie dla szeroko zakrojonego monitorowania pracowników. Chcąc zrozumieć nowy poziom ingerencji, jaki umożliwia algorytmiczna analityka danych, należy podkreślić jej cechy wyróżniające – tj. wszechobecność i interoperacyjność.

Systemy algorytmiczne przetwarzają informacje pochodzące z coraz bardziej zróżnicowanego zestawu technologii informacyjnych, umożliwiających bezprecedensowy dostęp do bardzo szczegółowych danych. Różne systemy przetwarzania danych są przy tym interoperacyjne, co pozwala łączyć zestawy danych i uzyskiwać nowe informacje na temat pracownika, które nie były widoczne, gdy poszczególne zestawy danych gromadzono i przechowywano oddzielnie. Wreszcie, coraz potężniejsza infrastruktura obliczeniowa jest w stanie przetwarzać dane w czasie rzeczywistym, przewidywać przyszłość za pomocą analityki predykcyjnej i realizować niezliczone zadania dzięki rozwijającej się sztucznej inteligencji. Dla przykładu w 2020 roku firma Amazon, wykorzystując technologię opartą na analityce danych, przyjęła do pracy 1700 pracowników w ciągu jednego dnia (Economist 2020b), a firma Serco skróciła czas potrzebny na zatrudnienie pracownika z czterech tygodni do czterech dni (Economist 2020a).

Kryzys związany z pandemią COVID-19 znacząco przyspieszył wdrożenie inteligentnych systemów do rekrutacji, oceny czy monitorowania pracowników (Suder, Siibak 2021). Zmiana ta nie powinna mieć charakteru tymczasowego. Badania przeprowadzone m.in. przez Institute for the Future of Work sugerują, że sztuczna inteligencja docelowo będzie stanowiła istotny komponent trzech głównych obszarów zadaniowych współczesnego modelu zarządzania zasobami ludzkimi – tj. zatrudniania, bieżącego zarządzania i oceny wydajności (Simons 2020).

Entuzjaści technik algorytmicznych twierdzą, że zasilana przez sztuczną inteligencję eksploracja danych – obok niezaprzeczalnych korzyści, takich jak m.in. poprawa wydajności, bezpieczeństwa i higieny pracy czy zdolności do wykrywania nieuczciwych zachowań – zapewnia również większą obiektywność w podejmowaniu decyzji personalnych. Zakłada się, że usunięcie człowieka z procesu decyzyjnego wyeliminuje

również ludzkie uprzedzenia czy funkcjonujące w świadomości społecznej stereotypy. Choć brzmi obiecująco, jest to tylko połowa prawdy. Od dawna w literaturze przedmiotu podkreśla się bowiem, że podejmowanie stronnicych i dyskryminacyjnych decyzji jest jednym z największych wyzwań sztucznej inteligencji. W praktyce algorytm jest bowiem tak dobry, jak dane historyczne, na których operuje (Kamiran 2013; Žliobaitė 2017).

Zagrożenia związane z dyskryminacją algorytmiczną generowane są na różnych etapach: od programowania/budowania, szkolenia i testowania do eksploatacji algorytmów. Dyskryminacja może wystąpić zarówno w relatywnie prostych algorytmach opartych na regułach (np. ludzkie uprzedzenia czy stereotypy mogą wpływać na parametry i reguły algorytmów na etapie opracowywania algorytmu, generując tym samym dyskryminujący wynik), jak i algorytmach uczenia maszynowego, które opierając się na korelacjach zidentyfikowanych w dużych ilościach danych społecznych i osobowych, mogą odtwarzać wzorce nierówności z przeszłości (Surden 2019, s. 1310–1320). Dyskryminacyjny potencjał algorytmów jest przy tym wysoce problematyczny niezależnie od tego, czy algorytmy są wykorzystywane do w pełni zautomatyzowanego podejmowania decyzji, czy tylko do wspierania procesu podejmowania decyzji przez człowieka.

W porównaniu z tradycyjnymi formami dyskryminacji, dyskryminacja algorytmiczna, ze względu na złożoną, a przez to nieprzejrzystą konstrukcję algorytmów jest bardziej abstrakcyjna i nieintuicyjna, a przez to trudna do wykrycia i udowodnienia¹. Nierzadko nie mieści się ona w kategoriach prawnych, które tradycyjnie definiują zakazaną dyskryminację. Jednocześnie ukrywa się ona za technologicznymi interfejsami, indukującymi dystans pomiędzy stosownymi procesami decyzyjnymi, a jednostkami pozbawionymi adekwatnych kompetencji cyfrowych do ich zrozumienia. Dystans ten manifestuje się w asymetrii informacyjnej oraz negocjacyjnej stron (Hildebrandt 2015; Koivisto 2020).

Nie bez powodu algorytmiczne podejmowanie decyzji zostało opisane jako nowa „granica regulacyjna” (m.in. dla ustawodawstwa antydyskryminacyjnego) (Hacker, Petkova 2017), a zaprezentowany w kwietniu 2021 roku przez Komisję Europejską projekt przepisów mających uregulować tę sferę technologiczną – Akt o sztucznej inteligencji (AI Act), wyraźnie uznaje za „kontekst wysokiego ryzyka” wdrażanie systemów sztucznej inteligencji do zadań związanych z zarządzaniem algorytmicznym (tj. zarządzaniem pracownikami i dostępem do samozatrudnienia, w szczególności do rekrutacji i wyboru kandydatów, do podejmowania decyzji o awansie i rozwiązaniu stosunku pracy oraz do przydzielania zadań, monitorowania lub oceny osób pozostających w umownych stosunkach pracy)².

¹ Algorytmy wykorzystywane do celów komercyjnych są przy tym często zastrzeżone, tj. chronione prawem własności intelektualnej i tajemnicą handlową.

² Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (AI Act) and Amending Certain Union Legislative Acts, COM(2021) 206 final, s. 4.

Niniejszy artykuł stanowi próbę zilustrowania kluczowych problemów i wyzwań w kontekście zatrudnienia, związanych z szeroko udokumentowaną na świecie zdolnością narzędzi sztucznej inteligencji do generowania dyskryminujących wyników w odniesieniu do szeregu cech chronionych. W niniejszym artykule za przedmiot dociekań przyjęto unijne ustawodawstwo antydyskryminacyjne, ustanawiające minimalne gwarancje ochrony przed dyskryminacją na poziomie krajowym³. Punktem wyjścia stosownych rozważań jest próba zrozumienia, z jakim rodzajem agencyjności bądź sprawczości mamy do czynienia w przypadku algorytmów. W celu przybliżenia mechanizmów i problematycznych aspektów zarządzania algorytmicznym artykuł łączy przy tym dwa nurty nauk prawnych: prawo nowych technologii oraz prawo pracy.

W stronę lepszego zrozumienia mechanizmu dyskryminacji algorytmicznej

Termin „sztuczna inteligencja” jest trudny do jednoznacznego zdefiniowania i bywa używany zamiennie z innymi terminami, takimi jak uczenie maszynowe, algorytmiczne podejmowanie decyzji czy zautomatyzowane podejmowanie decyzji (Komisja Europejska 2020). W dużym uproszczeniu pojęcie to można zdefiniować jako obejmujące zestaw technologii mających na celu „uczynienie komputerów zdolnymi do robienia rzeczy, jakie potrafią [ludzkie] umysły” (Boden 2018, s. 6). Aby to osiągnąć, sztuczna inteligencja potrzebuje dwóch elementów: kodu/zestawu reguł zwanych algorytmami (instrukcji matematycznych, których celem jest rozwiązywanie problemów, odpowiadanie na pytania lub wykonywanie określonych zadań), oraz danych, czyli zestawu zmiennych wejściowych. Jedną ze złożonych klas algorytmów jest tzw. uczenie maszynowe, czyli „zdolność [maszyn] do zdobywania własnej wiedzy poprzez wydobywanie wzorców z surowych danych” (Goodfellow, Bengio, Courville 2016, s. 2–3). Systemy uczenia maszynowego nie są przy tym bezpośrednio zaprogramowane do przestrzegania zestawu reguł, ale uczą się, jak dokładnie oszacować wynik na podstawie ilościowych zestawów danych, które wykorzystywane są do szkolenia i oceny wydajności modelu w czasie. To właśnie owa zdolność uczenia się i odkrywania korelacji w zbiorze danych znacząco zwiększa ryzyko powielania istniejących wzorców nierówności, nierzadko w sposób niezamierzony przez ich projektantów.

Systemy uczenia maszynowego wykorzystują bowiem dane o przeszłości do podejmowania decyzji, które kształtują przyszłość. Prognozy uczenia maszynowego nie są jednak przewidywaniami dotyczącymi konkretnej osoby, ale grup, które dzielą pewne cechy (np. płeć, wiek) lub wzorce zachowań. Jak wskazuje Mireille Hildebrandt (2015),

³ Notabene, zdecydowana większość dostępnych badań nad zagadnieniem algorytmów i dyskryminacji w sferze zatrudnienia nadal dotyczy głównie Stanów Zjednoczonych. Największym echem odbiły się przy tym, również na starym kontynencie, w szczególności rozważania zaprezentowane przez: Barocas, Selbst 2016; Kim 2016.

podstawową funkcją analityki algorytmicznej jest sortowanie i kategoryzowanie ludzi oraz tworzenie kolektywów w oparciu o różnice. Kultura algorytmów opiera się zatem na redukcji, uproszczeniu i budowaniu modeli. W ten sposób niepoliczalna w wielu wymiarach istota ludzka staje się „osobą policzalną” (Szpunar 2019, s. 25). Problemem konstytutywnym algorytmicznego podejmowania decyzji jest zatem to, że opierając się na poszukiwaniu stosownych podobieństw i różnic, *per se* generuje nierówności.

W literaturze przedmiotu zaproponowano przy tym kilka typologii opisujących, w jaki sposób kod/reguła lub dane mogą być stronicze, zatruwając tym samym operacje algorytmiczne dyskryminacją (Hacker 2018, s. 1146– 1148; Kleinberg 2019, s. 21–23; Barocas, Selbst 2016, s. 677–691⁴). Z perspektywy prawa antidyskryminacyjnego, mechanizmy, przez które dyskryminacja może być inkorporowana do algorytmów, mogą być w dużym uproszczeniu sklasyfikowane wzdłuż dwóch głównych osi. Dyskryminacja algorytmiczna może zarówno wynikać z hierarchizujących przekonań, stereotypów, jak i utrwalać strukturalne, zinstytucjonalizowane wzorce nierówności. Co istotne, wykluczenie kategorii chronionych z danych dostępnych dla algorytmów uczenia maszynowego nie zapobiega dyskryminacyjnym wynikom, ponieważ maszyny nierzadko używają pozornie neutralnych danych (np. miejsce zamieszkania) jako przybliżenia dla tych kategorii (pochodzenie etniczne, rasa) (Williams, Brooks, Shmargad 2018). Dla przykładu, algorytm uczenia maszynowego zaprojektowany w celu podejmowania decyzji o awansie w miejscu pracy i opierający się na danych osób, które w przeszłości odniosły „sukces”, może być ślepy na płeć kandydatów, a mimo to być dyskryminujący wobec kobiet, jeśli zostanie poproszony o wybranie kandydatów na podstawie oceny ich wydajności w pracy, która obejmowałaby średni czas pracy⁵. Istnieje zatem duże prawdopodobieństwo, że w niekorzystnej sytuacji znalazłyby się kobiety, ponieważ to na nich, zgodnie z tradycyjnym podziałem ról w społeczeństwie, spoczywa nieproporcjonalnie duży ciężar opieki nad dziećmi, a tym samym statystycznie kobiety pracują mniej godzin lub częściej w niepełnym wymiarze czasu pracy (Küllmann 2018)⁶.

⁴ Barocas i Selbst, notabene rozpatrując problem dyskryminacji algorytmicznej w zatrudnieniu w kontekście amerykańskiego ustawodawstwa antidyskryminacyjnego, proponują pięciopoziomową taksonomię źródeł dyskryminacji. Pierwszy poziom odpowiada temu, co autorzy nazywają „definicją problemu”, tj. konstrukcją „zmiennych docelowych”, które z kolei dzielą się na „etykiety klasowe”. Drugim poziomem stanowią dane treningowe wykorzystywane w procesie uczenia, które mogą być tendencyjne albo ze względu na sposób, w jaki zostały zebrane, albo dlatego, że same są tendencyjne i wykorzystywane do uczenia algorytmu. Trzecim poziomem jest wybór cech istotnych dla modelu, czyli tego, jakie atrybuty powinny być uznane za istotne w modelu algorytmicznym. Czwartym jest „dyskryminacja pośrednia”, wynikająca z tzw. „nadmiarowego kodowania”, czyli „przypadków, w których przynależność do klasy chronionej jest zakodowana w innych danych”, uznawanych za skorelowane z wynikiem przez model algorytmiczny. Piątym i ostatnim poziomem jest maskowanie celowej dyskryminacji, którą programiści mogą ukryć w architekturze algorytmu.

⁵ Na przykład w 2018 roku Amazon zrezygnował z algorytmu przesiewowego, który „karał” aplikacje na dane stanowisko / wnioski o pracę zawierające słowo „kobieta” (por. Dastin 2018).

⁶ Analogicznie, wynagrodzenie kierowców platform cyfrowych typu Uber zależy od ich oceny przez algorytm, który oblicza wynik za pomocą takich cech, jak wskaźnik dostępności i oceny klientów. Wykorzystanie tych cech wprowadza do procesu oceny zarówno stereotypy, jak i nierówności strukturalne.

Nie bez powodu algorytmy uczenia maszynowego, a w szczególności algorytmy głębokiego uczenia, często charakteryzowane są jako tzw. „czarne skrzynki”: ich wewnętrzne działanie nie jest zrozumiałe dla laików, a czasami nawet dla ekspertów, co utrudnia prześledzenie i wyizolowanie źródła dyskryminacji algorytmicznej (Leese 2014). Co więcej, nawet jeśli stosowny kod/reguła/dane zostaną zidentyfikowane, wykorzystanie zmiennych i kategoryzacja danych w algorytmach uczenia maszynowego podlegają ciągłej ewolucji w miarę uczenia się modelu. Kategorie te nie są statyczne, co w praktyce oznacza, że wszelkie dowody dyskryminacji algorytmicznej mogą zniknąć po zmutowaniu algorytmu.

W świetle powyższych uwag nie powinno zatem ulegać wątpliwości to, że przejście na zarządzanie algorytmiczne – niezależnie od tego, czy systemy algorytmiczne są wykorzystywane do podejmowania decyzji bez pomocy człowieka (w pełni zautomatyzowane podejmowanie decyzji), czy tylko do informowania i kształtowania ludzkiego procesu decyzyjnego – stanowi istotną zmianę jakościową w stosunkach i realiach pracy (Bąba 2020). Sposoby wdrażania systemów algorytmicznych przez przedsiębiorstwa tworzą środowisko, w którym normalizuje się wysoki poziom monitorowania i zautomatyzowanego podejmowania decyzji dotyczących podstawowych aspektów pracy, takich jak dostęp do zatrudnienia, awans czy wynagrodzenie. Algorytmizacja funkcji zarządczych zmienia przy tym medium, przez które dystrybuowana jest niejako nierówność. Stosowna zmiana medium, indukując kolejny poziom złożoności w stosunkach pracy, stawia pod znakiem zapytania możliwość przeciwdziałania i zadośćuczynienia dyskryminacji algorytmicznej w ramach tradycyjnego instrumentarium antydyskryminacyjnego, o czym szerzej w dalszej części tekstu.

Dyskryminacja algorytmiczna w świetle europejskiego ustawodawstwa antydyskryminacyjnego

Niniejsze rozważania koncentrować się będą na dwóch głównych narzędziach pojęciowych zawartych w prawie UE dotyczącym niedyskryminacji – tj. dyskryminacji bezpośredniej i dyskryminacji pośredniej. Kluczowe wyzwanie związane z kwalifikacją prawną dyskryminacyjnych procesów zarządzania algorytmicznego zasadniczo wydaje się bowiem polegać na tym, że w zależności od kontekstu i od tego, czy uwaga skupiona jest na samej operacji algorytmicznej, czy na włączeniu jej wyników do procesu decyzyjnego, dyskryminacja może być rozumiana jako bezpośrednia lub pośrednia. W praktyce, biorąc pod uwagę swoistego rodzaju fragmentację algorytmicznych systemów decyzyjnych, a przede wszystkim złożoność relacji człowiek–maszyna, rozróżnienie między tymi dwoma

Powołując się na przykład Ubera, Küllman (2018) dowodzi, jak stereotypy dotyczące niezetelności kobiet mogą wpływać na oceny klientów i w konsekwencji na ogólną ocenę kierowców, a tym samym na obliczanie ich wynagrodzenia, stawiając w niekorzystnej sytuacji kobiety.

pojęciami nie zawsze będzie jednoznaczne. Wydaje się, że ową trudność w pełni ilustruje przyjęta przez ustawodawcę europejskiego konstrukcja prawna zakazu dyskryminacji.

Dyskryminacja bezpośrednia

Dyskryminacja bezpośrednia zdefiniowana jest w prawie UE jako sytuacja, kiedy „osobę traktuje się mniej korzystnie/przychylnie (element szkody, *harm element*), niż traktuje się, traktowano lub traktowano by inną osobę w porównywalnej sytuacji”, zaś przyczyną takiego stanu rzeczy jest konkretna właściwość tej osoby, kwalifikowana jako „cecha chroniona” (element kauzalny, *causation element*)⁷. Dyskryminacja bezpośrednia ma więc charakter jawny. Istotą dyskryminacji bezpośredniej jest zaś różnica w traktowaniu, jakiej doświadcza dana osoba (Craig, de Búrca 2011, s. 896). Posługując się pojęciem gorszego traktowania w definiowaniu dyskryminacji bezpośredniej, ustawodawca europejski wprowadza przy tym element relatywny, wymuszający dokonywanie porównania z położeniem innych osób (porównywanego podmiotu, *comparator*). Porównanie może przy tym dotyczyć sytuacji aktualnej, przeszłej lub nawet hipotetycznej. Udowodnienie przypadku dyskryminacji bezpośredniej nie wymaga jednak ani wykazania, że sprawca był świadomy dyskryminacji, ani że miał zamiar dyskryminować (Maliszewska-Nienartowicz 2015, s. 79).

Konstrukcja ta potencjalnie pozwala objąć pojęciem dyskryminacji bezpośredniej szeroki zakres sytuacji, w których cechy chronione byłyby wykorzystywane jako istotne zmienne w modelu algorytmicznym, nawet jeśli intencją ich twórców czy użytkowników (pracodawców) nie była dyskryminacja. W kontekście uczenia maszynowego walor aplikacyjny stosownej koncepcji będzie jednak znikomy. W praktyce koncepcja ta obejmie bowiem jedynie sytuacje, kiedy modele (jakikolwiek element reguły lub kodu algorytmicznego) nie są neutralne w stosunku do chronionej cechy. Może to mieć miejsce na przykład w sytuacji, gdy chroniona cecha jest bezpośrednio wprowadzona do modelu algorytmicznego jako istotna zmienna i traktowana jako czynnik negatywny. Innymi słowy, jeśli osoba podejmująca decyzję wyraźnie konstruuje zbiór danych w taki sposób, że pewne grupy (np. kobiety czy osoby powyżej określonego wieku) są niedoreprezentowane, lub wybiera „gorszych” kandydatów z jednej chronionej grupy do celów szkoleniowych, model algorytmiczny będzie odzwierciedlał bezpośrednią dyskryminację.

Co istotne, bezpośrednie wprowadzenie dyskryminacji do modelu algorytmicznego prawdopodobnie zmniejszy jego wartość predykcyjną, co stanowi istotny czynnik

⁷ Zob. art. 2 ust. 2 lit. a dyrektywy Rady 2000/43/WE z dnia 29 czerwca 2000 roku, wprowadzającej w życie zasadę równego traktowania osób bez względu na pochodzenie rasowe lub etniczne (Dz.Urz. UE L 2000, nr 180/22); art. 2 ust. 2 lit. a dyrektywy Rady 2000/78/WE z dnia 27 listopada 2000 roku ustanawiającej ogólne warunki ramowe równego traktowania w zakresie zatrudnienia i pracy (Dz.Urz. UE L 2000, nr 303/16); art. 2 ust. 1 lit. a dyrektywy 2006/54/WE Parlamentu Europejskiego i Rady z dnia 5 lipca 2006 roku w sprawie wprowadzenia w życie zasady równości szans oraz równego traktowania kobiet i mężczyzn w dziedzinie zatrudnienia i pracy (wersja preredagowana) (Dz.Urz. UE L 2006, nr 204/23).

zniechęcający (Hacker 2018, s. 1152). Bardziej prawdopodobnym scenariuszem, w ramach którego może dojść do dyskryminacji bezpośredniej, jest w związku z tym sytuacja, gdy algorytm bazuje na tzw. parametrze zastępczym (*proxy*), bezpośrednio skorelowanym z cechą chronioną. Dla przykładu, miałyby to miejsce, gdyby algorytm obniżał ocenę każdej rozmowy o pracę przeprowadzonej zdalnie z osobą o ciemniejszej karnacji. Dyskryminacja bezpośrednia wystąpiłaby w tym kontekście albo dlatego, że czynnikiem dyskryminującym jest sama cecha chroniona (pochodzenie etniczne), albo dlatego, że istnieje zależność między osobami pokrzywdzonymi przez dyskryminujący algorytm a osobami posiadającymi cechę chronioną (Kelly-Lyth 2021, s. 8).

W świetle powyższych uwag, w literaturze przedmiotu zdecydowanie większy potencjał aplikacyjny w przypadku dyskryminacji algorytmicznej przypisuje się rozwiniętej przez Trybunał Sprawiedliwości Unii Europejskiej (dalej: Trybunał, TSUE) koncepcji dyskryminacji przez asocjacje i dyskryminacji percepcyjnej. Chodzi o rozszerzenie koncepcji dyskryminacji bezpośredniej odpowiednio na sytuacje, kiedy dana osoba jest traktowana niekorzystnie, ze względu na więź z osobą posiadającą cechę prawnie chronioną (wyrok TSUE z dnia 17 lipca 2008 roku, w sprawie Coleman przeciwko Attridge Law i Steve Law, C-303/06, ECLI:EU:C:2008:415; Naumann 2007, s. 286), bądź na sytuacje, gdy osoba dopuszczająca się dyskryminacji błędnie zakłada, że dana osoba posiada określoną cechę prawnie chronioną (Zwolińska 2010, s. 31). Stosowne rozszerzenie koncepcyjne jest niezwykle cenne w kontekście dyskryminacji algorytmicznej, ponieważ pozwala objąć dyskryminacją bezpośrednią niektóre przypadki dyskryminacji algorytmicznej na zasadzie przybliżenia i błędnej kategoryzacji. Przykładowo, dyskryminacja bezpośrednia mogłaby obejmować przypadki profilowania algorytmicznego, w którym użytkownicy, mimo że sami nie posiadają danej cechy, są zaliczani przez algorytm do kategorii chronionej. W praktyce takie błędy w klasyfikacji związane z wykorzystaniem danych zastępczych w profilowaniu algorytmicznym mogłyby zaistnieć w sytuacji, kiedy algorytm nie przedkładałby użytkownikowi serwisu społecznościowego (typu Facebook, LinkedIn) ofert pracy w danej branży, np. z powodu błędnej klasyfikacji jako osoby niepełnosprawnej, bazującej na danych behawioralnych (polubieniach czy częstotliwości odwiedzania stron związanych z tą tematyką).

Co istotne, w opinii Komisji Europejskiej zarówno dyrektywa o równości rasowej, jak i dyrektywa o równym traktowaniu „zakazują również sytuacji, w której osoba jest bezpośrednio dyskryminowana na podstawie błędnego postrzegania lub przypuszczenia dotyczącego cech chronionych” (Komisja Europejska 2014). Interpretacja ta nie była jednak dotychczas konsekwentnie stosowana przez Trybunał Sprawiedliwości⁸. Dla przykładu, w sprawie Karstena Kaltofta, opiekuna dziecięcego postrzeganego przez pracodawcę jako osoba niepełnosprawna z powodu otyłości i z tego powodu zwolnionego

⁸ TSUE do tej pory interpretował jedynie pojęcie dyskryminacji pośredniej jako obejmujące sytuacje, w których błędne wyobrażenia skutkują dyskryminacyjnym traktowaniem jednostek (wyrok TSUE z dnia 16 lipca 2015 roku, w sprawie CHEZ Razpredelenie Bulgaria AD przeciwko Komisja za zashtita ot diskriminatsia, C-83/14, ECLI:EU:C:2015:480).

z pracy (wyrok TSUE z dnia 18 grudnia 2014 roku, w sprawie Fag og Arbejde (FOA) przeciwko Kommunernes Landsforening (KL), C-354/13, ECLI:EU:C:2014:2463), Trybunał Sprawiedliwości zamiast uznać, że stosowne postrzeganie otyłości jako niepełnosprawności przez pracodawcę spowodowało odmienne traktowanie i dyskryminację, zobowiązał sąd krajowy do ustalenia, czy stosowne „trwałe osłabienie funkcji fizycznych, umysłowych lub psychicznych” miało miejsce, stwierdzając, że jest to warunek uznania otyłości za niepełnosprawność i ostatecznie stwierdzenia dyskryminacji ze względu na niepełnosprawność.

Restrykcyjna wykładnia koncepcji dyskryminacji bezpośredniej jest przy tym problematyczna w świetle algorytmicznych technik profilowania, które nierzadko koncentrują się wyłącznie na określonych zachowaniach, zainteresowaniach czy podobieństwach i na tej podstawie wyciągają wnioski na temat tożsamości czy szeroko rozumianej przynależności społecznej jednostek. W konsekwencji mimo że pojęcie dyskryminacji bezpośredniej nadaje się do uchwycenia pewnych dyskryminujących operacji algorytmicznych, jego praktyczne zastosowanie będzie prawdopodobnie bardzo ograniczone, w szczególności w świetle oczywistych trudności dowodowych. Sygnalizowany powyżej brak przejrzystości funkcjonowania algorytmu nie tylko utrudnia zrozumienie, czy parametry modelu algorytmicznego są neutralne wobec kategorii chronionych, ale również uniemożliwia zgromadzenie dowodów na to, jak algorytm potraktował/by osobę która nie podziela danej cechy chronionej. Unijne ustawodawstwo antydyskryminacyjne nie ustanawia przy tym „prawa pracownika, który twierdzi w sposób przekonujący, że spełnia warunki określone przez pracodawcę w ogłoszeniu o rekrutacji, ale jego kandydatura nie została uwzględniona, do uzyskania informacji na temat tego, czy w wyniku procedury rekrutacyjnej zatrudnił on innego kandydata” (wyrok TSUE z dnia 19 kwietnia 2012 roku, w sprawie Galina Meister przeciwko Speech Design Carrier Systems GmbH, C-415/10, ECLI:EU:C:2012:217). Ustalenie bezpośredniej dyskryminacji algorytmicznej, w szczególności na etapie rekrutacji pracownika, może być zatem skomplikowane ze względu na wspomniane wyżej przeszkody.

Dyskryminacja pośrednia

Częściowe odkodowanie algorytmicznej „czarnej skrzynki” z dyskryminacji wydaje się zapewniać unijna koncepcja dyskryminacji pośredniej. Ta ostatnia ma miejsce „w przypadku, gdy z pozoru neutralny przepis, kryterium lub praktyka stawiałaby osoby należące do jednej z [chronionych grup] w szczególnie niekorzystnej sytuacji w porównaniu do [innych osób], chyba że dany przepis, kryterium lub praktyka są obiektywnie uzasadnione zgodnym z prawem celem, a środki mające służyć osiągnięciu tego celu są właściwe i niezbędne”⁹. W grę wchodzi przy tym zarówno działanie celowe, jak również

⁹ Zob. art. 2 ust. 2 lit. b dyrektywy Rady 2000/43/WE z dnia 29 czerwca 2000 roku, wprowadzającej w życie zasadę równego traktowania osób bez względu na pochodzenie rasowe lub etniczne (Dz.Urz. UE L 2000, nr 180/22); art. 2 ust. 2 lit. b dyrektywy Rady 2000/78/WE z dnia 27 listopada 2000 roku

nieświadomy rezultat sposobu funkcjonowania pracodawcy. Istotne jest jednak, by stosowne przyczyny różnicowania miały charakter pozornie neutralny. Jednocześnie nie jest konieczne, aby faktycznie doszło do niekorzystnego ukształtowania położenia danej grupy – wystarczy, aby dana praktyka, kryterium lub przepis były w stanie wygenerować szczególnie niekorzystne położenie.

Trzy elementy konstrukcyjne koncepcji dyskryminacji pośredniej jawią się przy tym jako szczególnie istotne w przypadku dyskryminacji algorytmicznej. Po pierwsze, przesuwając akcent na pozorną neutralność kryterium czy praktyki, unijna koncepcja dyskryminacji pośredniej wydaje się odpowiednia do uchwycenia szerokiego spektrum pozornie neutralnych, ale w rzeczywistości dyskryminujących algorytmów, np. kiedy zmienne korelujące z grupą chronioną są wykorzystywane jako istotne cechy lub etykiety w algorytmie. Ponadto koncepcja ta – przenosząc punkt ciężkości analizy na skutki działania algorytmów, a nie na ich zasady, parametry czy treść – nie czyni koniecznym otwarcia algorytmicznej „czarnej skrzynki”, co zwiększa możliwość kwalifikacji określonych niekorzystnych skutków zarządzania algorytmicznego jako dyskryminujących. Od wnioskodawcy znajdującego się w niekorzystnej sytuacji nie będzie się bowiem wymagać wykazania, dlaczego algorytm ocenił go nisko; sam fakt niekorzystnej sytuacji powinien być wystarczający, aby wymagać stosownego uzasadnienia jego zastosowania. W końcu, skupienie koncepcji dyskryminacji pośredniej na grupie dotkniętej środkiem dyskryminującym, a nie na indywidualnym wnioskodawcy (McCrudden, Prechal 2009, s. 35), jako spójne z opisywanym wcześniej sposobem funkcjonowania algorytmów, powinno pomóc rzucić światło na nierzadko replikowaną przez algorytmy systemową i strukturalną dyskryminację. Trybunał Sprawiedliwości w swoim orzecznictwie nie wymaga przy tym precyzyjnego dowodu stopnia, w jakim dany środek wpływa na konkretną grupę osób (Maliszewska-Nienartowicz 2012, s. 253). W praktyce, wystarczająco statystyczna różnica między wynikami grupy lub algorytmiczna preferencja dla cechy, która jest mniej powszechna wśród chronionej grupy wnioskodawcy (Kelly-Lyth 2021, s. 8). Standard ten w przypadku algorytmów powinien nastęrczać większych trudności, ponieważ statystyka jest podstawą uczenia maszynowego, a stosowna dokładność predykcyjna jest zawsze mierzona (Shwartz, David 2014, s. 14).

Z pojęciem dyskryminacji pośredniej i jego zastosowaniem w praktyce wiążą się jednak również określone komplikacje. O ile stwierdzenie dyskryminacji bezpośredniej wyklucza uzasadnienie poza zamkniętą i ograniczoną listą wyjątków¹⁰, o tyle stwierdzenie dyskryminacji pośredniej *prima facie* otwiera szeroki wachlarz możliwych uzasadnień. Dyrektywy unijne wyraźnie stanowią bowiem, że przepis, kryterium lub praktyka nie są zakazane w sytuacji, gdy są obiektywnie uzasadnione zgodnym z prawem celem, a środki

ustanawiający ogólne warunki ramowe równego traktowania w zakresie zatrudnienia i pracy (Dz.Urz. UE L 2000, nr 303/16); art. 2 ust. 1 lit. b dyrektywy 2006/54/WE Parlamentu Europejskiego i Rady z dnia 5 lipca 2006 roku w sprawie wprowadzenia w życie zasady równości szans oraz równego traktowania kobiet i mężczyzn w dziedzinie zatrudnienia i pracy (wersja przededytowana) (Dz.Urz. UE L 2006, nr 204/23).

¹⁰ Zob. art. 4 dyrektywy 2000/43/WE; art. 4 i 6 dyrektywy 2000/78/WE; oraz art. 14 ust. 2 dyrektywy 2006/54/WE.

osiągnięcia tego celu są właściwe i niezbędne. Stosowny element konstrukcyjny otwiera pracodawcy możliwość powołania się na dowolne, zgodne z prawem uzasadnienie i poddanie go pod rozważę sądu.

Katalog celów zgodnych z prawem jest otwarty. W odniesieniu do środków stosowanych przez pracodawców Trybunał Sprawiedliwości wprost uznaje przy tym cele o ekonomicznym charakterze, jak np. efektywność przedsiębiorstwa czy sytuacja na rynku pracy (Maliszewska-Nienartowicz 2012, s. 394–395). W kontekście zarządzania algorytmicznego, pracodawca wydaje się mieć zatem do dyspozycji dość szeroki zakres możliwości eksplanacyjnych. Algorytmizacja wybranych funkcji zarządczych może być dyktowana koniecznością zapewnienia bezpiecznych i higienicznych warunków pracy, efektywnego wykorzystania czasu pracy czy optymalizacją i obiektywizacją procesów decyzyjnych dotyczących awansów, szkoleń, czy zwolnień. Zasadne wydaje się zatem ograniczenie przez TSUE kręgu celów uzasadniających dyskryminację pośrednią w kontekście zatrudnienia. Istotną rolę w tym zakresie wydaje się przy tym również odgrywać zasada proporcjonalności.

Co do zasady, jeżeli cel zostanie uznany za zgodny z prawem, przedmiotowy środek musi zostać uznany za odpowiedni i niezbędny (tj. zgodny z zasadą proporcjonalności). W przypadku algorytmów sądy prawdopodobnie skłonne będą uznać, że są one skutecznymi środkami do osiągnięcia celu, jakim jest szeroko rozumiana optymalizacja procesów zarządczych. Algorytmy są w rzeczywistości opracowywane właśnie po to, aby zapewnić poziom precyzji, szczegółowości i efektywności, którego ludzkie umysły nie są w stanie odtworzyć. W związku z tym wymogi dotyczące uzasadnionego celu i odpowiedniości/skuteczności będą prawdopodobnie relatywnie łatwo spełnione. Problematyczne natomiast powinno okazać się w praktyce wykazanie proporcjonalności *sensu stricto*, która pozwala na wyważenie korzyści uzyskiwanych dzięki utrzymaniu tak uzasadnionej praktyki z naruszeniem zakazu dyskryminowania.

Zgodnie z wymogiem niezbędności – nie może istnieć „żaden inny środek do osiągnięcia [tego samego] celu, który w mniejszym stopniu naruszałby prawo do niedyskryminacji” (EU Agency for Fundamental Rights, Council of Europe 2018, s. 45). Trybunał Sprawiedliwości i sądy krajowe powinny zatem być mniej skłonne do zaakceptowania szerokich uzasadnień ostatniej części testu proporcjonalności. Wykorzystanie stronniczego algorytmu zasadniczo powinno być uznane za nieproporcjonalne, jeśli równie zadowalające i mniej dyskryminujące alternatywy okażą się łatwo dostępne. W związku z tym pracodawca będzie musiał podjąć rozsądne wysiłki w celu zapewnienia, że wdrażany algorytm jest najmniej dyskryminującym środkiem do osiągnięcia zamierzonego celu. W praktyce można jednak wyobrazić sobie, że algorytm, który wykazuje pewne uprzedzenia, może być mniej dyskryminujący niż alternatywna opcja – tj. pracownik działu kadr czytający CV i decydujący o tym, z kim przeprowadzić rozmowę kwalifikacyjną. Innymi słowy, chcąc utrzymać przewagę konkurencyjną, którą w świecie rządzonej przez informacje i dane niezaprzeczalnie generuje algorytmizacja funkcji zarządczych, pracodawca może dążyć do wykazania, że wykorzystanie algorytmu w rzeczywistości jest najmniej dyskryminującą dostępną opcją. W tym sensie, w przypadku dyskryminacji algorytmicznej, przesłanka

proporcjonalności prawdopodobnie przerodzi się w pytanie o dopuszczalny w sferze zatrudnienia kompromis między szeroko rozumianą efektywnością biznesową a zasadą niedyskryminacji. W świetle powyższego, jedynie rygorystyczna ocena proporcjonalności wdrożonych przez pracodawców rozwiązań algorytmicznych może przyczynić się do zwiększenia skuteczności zakazu dyskryminacji pośredniej w praktyce.

Podsumowując: unijna koncepcja dyskryminacji pośredniej wydaje się wystarczająco pojemna do objęcia swoim zakresem szerokiego spektrum praktyk związanych z zarządzaniem algorytmicznym, niemniej jednak otwartość testu proporcjonalności stawia pod znakiem zapytania jej rzeczywisty potencjał osłabiania dyskryminacji algorytmicznej. Problematyczny okazać się może przy tym zarówno szeroki wachlarz dopuszczalnych uzasadnień (w praktyce może on rodzić negatywne implikacje dla pewności prawa, ponieważ ocena relewantności potencjalnych uzasadnień będzie dokonywana *post factum* przez sądy¹¹), jak i kryterium niezbędności, które rodzi pytania o dopuszczalny zakres kompromisu między modernizacją i optymalizacją procesów zarządczych a zasadą niedyskryminacji. Stosowne mankamenty konstrukcyjne obecnie wydają się wyłączać możliwość ustanowienia minimalnych standardów ochrony w odniesieniu do dyskryminacji algorytmicznej na poziomie ponadnarodowym, a w konsekwencji czynią pilną dyskusję naukową dotyczącą konieczności wypracowania efektywnych odpowiedzi regulacyjnych na poziomie krajowym¹².

Podsumowanie

Jak wykazała przeprowadzona rudymentarna analiza najważniejszych komponentów konstrukcyjnych ustanowionych przez prawodawcę unijnego podstawowych form dyskryminacji, unijne prawo antydyskryminacyjne oferuje pewne możliwości przeciwdziałania negatywnym skutkom dyskryminacji algorytmicznej w sferze zatrudnienia. W świetle braku skontekstualizowanego/systemowego podejścia do omawianego problemu na poziomie krajowym, stosowny potencjał ochronny musiałby jednak zostać potwierdzony

¹¹ Dominujący w prawie UE model niedyskryminacji jest systemem kontrydiktoryjnym *ex post*, opartym na indywidualnych sporach sądowych. Oznacza to, że ciężar wykrycia dyskryminacji, wszczęcia postępowania i wniesienia sprawy do sądu spoczywa na ofierze dyskryminacji (Craig, de Búrca 2011, s. 918).

¹² Obecnie jedynie w Hiszpanii zdecydowano się na wprowadzenie przepisów odnoszących się *explicite* do AI. Królewski dekret z mocą ustawy 9/2021 z dnia 11 maja 2021 roku zmienił Statut Pracowników, wprowadzając nowy obowiązek informowania przedstawicieli prawnych pracowników o AI i o stosowaniu algorytmów, ustanawiając, że rady pracowników (lub delegaci personelu, jeśli ma to zastosowanie w przedsiębiorstwach zatrudniających mniej niż 50 pracowników) muszą być informowane o: „parametrach, zasadach i instrukcjach, na których opierają się algorytmy lub systemy sztucznej inteligencji, które mają wpływ na podejmowanie decyzji dotyczących/wpływających na warunki pracy, dostęp i utrzymanie zatrudnienia, w tym profilowanie”. Niedopełnienie stosownego obowiązku informacyjnego stanowi poważne naruszenie i może prowadzić do nałożenia przez Inspekcję Pracy kary pieniężnej w wysokości od 750 do 7500 euro.

przez Trybunał Sprawiedliwości, przy czym zadanie to niewątpliwie będzie wymagało pewnej kreatywności interpretacyjnej w celu dostosowania konstrukcji normatywnej zakazu dyskryminacji do specyfiki operacyjnej algorytmów. Obecnie, w świetle braku takich rozwiązań, coraz powszechniejsze stosowanie algorytmów w sferze zatrudnienia zdaje się przenosić odpowiedzialność „w chmurę”.

Swoistego rodzaju fragmentacja łańcucha podejmowania decyzji dotyczących algorytmów stawia pod znakiem zapytania to, kto (twórca czy użytkownik algorytmu – pracodawca), w jakim zakresie, i na jakich zasadach powinien zostać pociągnięty do odpowiedzialności w przypadku dyskryminacji wynikającej z wykorzystania algorytmów. Jednocześnie, na poziomie egzekwowania prawa, wciąż silne poleganie na zindywidualizowanym podejściu opartym na prawach jednostki generuje wiele trudności związanych z dostępem do informacji i skuteczną ochroną sądową. Istnieje zatem wyraźna potrzeba zastanowienia się w dalszych badaniach nad tym, jak skutecznie rozwiązać te problemy na poziomie krajowym, a także nad tym, jakie mechanizmy nadzorcze można by wprowadzić, aby zmniejszyć obciążenia faktycznie ponoszone przez pracowników dotkniętych dyskryminacją algorytmiczną.

Istotny potencjał w zakresie przewyżczenia opisanych w niniejszym artykule trudności natury praktycznej wydaje się przy tym posiadać powiązanie konstrukcji prawa antydyskryminacyjnego z nowymi narzędziami egzekwowania prawa inkorporowanymi do rozporządzenia Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 roku 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) (Dz.Urz. UE L 2016, nr 119/1; dalej: RODO). Akt ten przewiduje bowiem nie tylko prawo dostępu do danych (art. 15) czy prawo do informacji o zautomatyzowanym podejmowaniu decyzji, w tym o profilowaniu (art. 13–15), ale również wprowadza obowiązek oceny skutków dla ochrony danych, w odniesieniu do operacji przetwarzania niosących wysokie ryzyko naruszenia praw lub wolności osób fizycznych (art. 35). Przesuwając punkt ciężkości z mechanizmów *ex post* na zasady *ex ante*, mające na celu zapobieganie nieuczciwemu przetwarzaniu danych, RODO pozwala zatem wypracować stosowne mechanizmy egzekwowania równego traktowania już w fazie projektowania algorytmu. Starannie przemyślany mariaż koncepcji prawa antydyskryminacyjnego i instrumentów prawnych z zakresu ochrony danych osobowych na poziomie ustawodawstwa krajowego mógłby w tym sensie stanowić obiecującą metodę walki z dyskryminacją algorytmiczną w sferze zatrudnienia.

Bibliografia

- Ajunwa I., Greene D. (2018) *Platforms at Work: Automated Hiring Platforms and Other New Intermediaries in the Organization of Work*, „Research in the Sociology of Work”, No. 33.
- Barocas S., Selbst A.D. (2016) *Big Data's Disparate Impact*, „California Law Review”, Vol. 104, No. 671.

- Bąba M. (2020) *Algorytmy – nowy wymiar nadzoru i kontroli nad świadczącym pracę*, „Praca i Zabezpieczenie Społeczne”, nr 3.
- Boden M.A. (2018) *Artificial Intelligence: A Very Short Introduction*, Oxford.
- Craig P., de Búrca G. (2011) *EU Law*, Oxford.
- Dastin J. (2018) *Amazon Scraps Secret AI Recruiting Tool that Showed Bias Against Women*, <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secretairecruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>.
- Economist (2020a) *How Covid-19 Is Driving Public-Sector Innovation*, <https://www.economist.com/britain/2020/04/03/how-covid-19-is-driving-public-sector-innovation>.
- Economist (2020b) *The Coronavirus Crisis Thrusts Corporate HR Chiefs Into the Spotlight*, <https://www.economist.com/business/2020/03/26/the-coronavirus-crisis-thrusts-corporate-hr-chiefs-into-the-spotlight>.
- EU Agency for Fundamental Rights, Council of Europe (2018) *Handbook on European non-discrimination law*, <https://fra.europa.eu/en/publication/2018/handbook-european-non-discrimination-law-2018-edition>.
- Goodfellow I., Bengio Y., Courville A. (2016) *Deep Learning*. The MIT Press.
- Hacker P. (2018) *Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies against Algorithmic Discrimination under EU Law*, „Common Market Law Review”, Vol. 55, No. 4.
- Hacker P., Petkova B. (2017) *Reining in the Big Promise of Big Data: Transparency, Inequality, and New Regulatory Frontiers*, „Northwestern Journal of Technology and Intellectual Property”, No. 15.
- Hildebrandt M. (2015) *Smart Technologies and the End(s) of Law: Novel entanglements of law and technology*, Cheltenham, UK.
- Kamiran F. et al. (2013) *Quantifying Explainable Discrimination and Removing Illegal Discrimination in Automated Decision Making*, „Knowledge & Information Systems”, No. 35.
- Kelly-Lyth A. (2021) *Challenging Biased Hiring Algorithms*, „Oxford Journal of Legal Studies”, Vol. 41, Issue 4.
- Kim P.T. (2016–2017) *Data-Driven Discrimination at Work*, „William and Mary Law Review”, Vol. 58, Issue 3.
- Kleinberg J. et al. (2019) *Discrimination in the Age of Algorithms*, 25548 NBER „Journal of Legal Analysis”, Vol. 10.
- Koivisto I. (2020) *Thinking Inside the Box: The Promise and Boundaries of Transparency in Automated Decision-Making*, Working Paper EUI AEL 2020/01, San Domenico di Fiesole.
- Komisja Europejska (2014) Joint Report on the application of Council Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin (‘Racial Equality Directive’) and of Council Directive 2000/78/EC of 27 November 2000 establishing a general framework for equal treatment in employment and occupation (‘Employment Equality Directive’) COM(2014) 2 final (Brussels).
- Komisja Europejska (2020) *White Paper on Artificial Intelligence: A European Approach to Excellence and Trust*, COM(2020) 65 final.

- Kullmann M. (2018) *Platform Work, Algorithmic Decision-Making, and EU Gender Equality Law*, „The International Journal of Comparative Labour Law and Industrial Relations”, Vol. 34, No. 1.
- Leese M. (2014) *The New Profiling: Algorithms, Black Boxes, and the Failure of Anti-discriminatory Safeguards in the European Union*, „Security Dialogue”, Vol. 45.
- Maliszewska-Nienartowicz J. (2012) *Dyskryminacja pośrednia w prawie Unii Europejskiej*, Toruń.
- Maliszewska-Nienartowicz J. (2015) *Konstrukcja prawna zakazu dyskryminacji w prawie Unii Europejskiej* [w:] A. Szczerba-Zawada, A. Zawidzka-Łojek (red.), *Prawo antydyskryminacyjne Unii Europejskiej*, Warszawa.
- McCrudden Ch., Prechal S. (2009) *The Concepts of Equality and Non-Discrimination in Europe: A Practical Approach*, EC Directorate-General for Employment, Social Affairs and Equal Opportunities, Unit G. 2.
- Naumann E. (2007), *Dyskryminacja w prawie pracy*, „Monitor Prawa Pracy”, nr 6.
- Otto M. (2019) „Workforce Analytics” v *Fundamental Rights Protection in the EU in the Age of Big Data*, „Comparative Labor Law & Policy Journals”, Vol. 40, Issue 3.
- Shwartz S., David B. (2014) *Understanding Machine Learning. From Theory to Algorithms*, Cambridge.
- Simons J. (2020) *Equality Task Force: Machine learning case studies*, Institute for the Future of Work.
- Suder S., Siibak A. (2021) *Proportionate Response to a COVID-19 Threat? Use of Apps and Other Technologies for Monitoring Employees Under the EU Data Protection Framework*, „International Labour Review”, Vol. 161, Issue 2 (Special Issue: „COVID-19 and the World of Work (part II)”).
- Surden H. (2019) *Artificial Intelligence and Law: An Overview*, „Georgia State University Law Review”, Vol. 35, Issue 4.
- Szpunar M. (2019) *Kultura algorytmów*, Kraków.
- Williams B.A., Brooks C.F., Shmargad Y. (2018) *How Algorithms Discriminate Based on Data They Lack: Challenges, Solutions, and Policy Implications*, „Journal of Information Policy”, Vol. 8.
- Taylor F. (1972) *Zasady naukowego zarządzania* [w:] J. Kurnal (red.), *Twórcy naukowych podstaw organizacji*, Warszawa.
- Wood A.J. (2021) *Algorithmic Management: Consequences for Work Organisation and Working Conditions* European Commission, „JRC Working Papers Series on Labour, Education and Technology”, Vol. 7.
- Žliobaitė I. (2017) *Measuring Discrimination in Algorithmic Decision Making*, „Data Mining and Knowledge Discovery”, Vol. 31, Issue 4.
- Zwolińska A. (2010) *Zakaz dyskryminacji w zatrudnieniu*, „Serwis Prawno-Pracowniczy”, nr 15.

Orzecznictwo

- Wyrok TSUE z dnia 17 lipca 2008 roku w sprawie Coleman przeciwko Attridge Law i Steve Law, C-303/06, ECLI:EU:C:2008:415.
- Wyrok TSUE z dnia 19 kwietnia 2012 roku w sprawie Galina Meister przeciwko Speech Design Carrier Systems GmbH, C-415/10, ECLI:EU:C:2012:217.

Wyrok TSUE z dnia 18 grudnia 2014 roku w sprawie Fag og Arbejde (FOA) przeciwko Kommunernes Landsforening (KL), C-354/13, ECLI:EU:C:2014:2463.

Wyrok TSUE z dnia 16 lipca 2015 roku w sprawie CHEZ Razpredelenie Bulgaria AD przeciwko Komisija za zashtita ot diskriminatsia, C-83/14, ECLI:EU:C:2015:480.

Akty prawa europejskiego

Dyrektywa Rady 2000/43/WE z dnia 29 czerwca 2000 roku wprowadzająca w życie zasadę równego traktowania osób bez względu na pochodzenie rasowe lub etniczne, Dz.Urz. UE L 2000, nr 180/22.

Dyrektywa Rady 2000/78/WE z dnia 27 listopada 2000 roku ustanawiająca ogólne warunki ramowe równego traktowania w zakresie zatrudnienia i pracy, Dz.Urz. UE L 2000, nr 303/16.

Dyrektywa 2006/54/WE Parlamentu Europejskiego i Rady z dnia 5 lipca 2006 roku w sprawie wprowadzenia w życie zasady równości szans oraz równego traktowania kobiet i mężczyzn w dziedzinie zatrudnienia i pracy (wersja przededagowana), Dz.Urz. UE L 2006, nr 204/23.

Rozporządzenia Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 roku 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), Dz.Urz. UE L 2016, nr 119/1.

Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (AI Act) and Amending Certain Union Legislative Acts, COM(2021) 206 final.