Zdrowie Publiczne i Zarządzanie 2017; 15 (4): 306–316
www.ejournals.eu/Zdrowie-Publiczne-i-Zarzadzanie, doi:10.4467/20842627OZ.17.033.8287

Zagadnienie stosowania samouczących się programów o charakterze sztucznych inteligencji w ochronie zdrowia

Krzysztof Kaźmierczak

Katedra Prawa Międzynarodowego i Stosunków Międzynarodowych, Wydział Prawa i Administracji, Uniwersytet Łódzki

Adres do korespondencji: Krzysztof Kaźmierczak, Katedra Prawa Międzynarodowego i Stosunków Międzynarodowych, Wydział Prawa i Administracji, Uniwersytet Łódzki, ul. Kopcińskiego 8/12, 90-232 Łódź, krzysztof.w.kazmierczakgmail.com

Abstract

An approach for using self-teaching algorithms of artificial intelligences character in health care

Recent developments in information technologies allow an increasing delegation of certain tasks within the area of medicine to self-teaching algorithms, commonly referred to as artificial intelligences. Such programs may advise and provide solutions in various aspects of personalized medical services to be used by practitioners. Large scale usage of data and creation of complicated pathways for decision making may result in opacity and creation of "black boxes" which cannot be effectively supervised by users. This paper aims to position certain aspects of such algorithms within the prescriptive map of ethical debate on algorithms, their accountability and to briefly describe the existing legal solutions which may aid in dealing with such issues.

Key words: **ethics of algorithms, big data, medical devices, profiling**

Słowa kluczowe: **etyka algorytmów, big data, narzędzia medyczne, profilowanie**

Wstęp

Personalizacja medycyny – dostosowanie określonych terapii do konkretnej osoby, niejednokrotnie wskazywana jest jako jeden z nadchodzących przełomów w zakresie ochrony zdrowia[1]. Oczywiste wydaje się stwierdzenie, że każdy pacjent jest inny, leczenie zaś powinno być dostosowane w możliwie najwyższym stopniu do konkretnej osoby. Ustalenie łączników pomiędzy cechami biologicznymi pacjenta a różnymi rezultatami proponowanego leczenia pozwala określić, w jaki konkretnie sposób dany pacjent będzie reagował na określoną formę leczenia. Dzięki takim informacjom możliwe jest zarówno bardziej efektywne świadczenie usług medycznych, jak i zmniejszenie kosztów i ryzyka związanych ze zbędnymi interwencjami medycznymi poprzez wskazanie faktycznych zależności[2]. Podobnie uwzględnienie szczegółowej charakterystyki konkretnej osoby może pozwolić na przyspieszenie i usprawnienie badań prowadzonych nad skutkami stosowania produktów leczniczych, eliminując niekorzystne działanie ich szerokiego badania na wybranych grupach, utrudniające ustalenie ich wpływu na określone grupy[3]. Aktualna forma dostosowania świadczeń medycznych do osoby opiera się w dużej mierze na wyraźnych i oczywistych relacjach, na podstawie których dana osoba przyporządkowana zostaje do określonych kategorii, utworzonych według wcześniej przeprowadzonych badań o charakterze statystycznym – poprzez przyporządkowanie ich do określonej grupy, dla której określone są dane wskazania bądź przeciwwskazania [4]. Taka forma personalizacji często nie uwzględnia bardziej skomplikowanych zależności pomiędzy aspektami mogącymi mieć wpływ na daną jednostkę i na skuteczność terapii ze względu na ograniczone możliwości dokonania analizy konkretnej sytuacji przez człowieka i tym samym ograniczenie liczby analizowanych w każdym przypadku czynników.

Rozwiązaniem pozwalającym poszerzyć zakres personalizacji medycyny jest wykorzystanie komputerowych programów o charakterze tak zwanej sztucznej inteligencji do podejmowania decyzji związanych z ochroną zdrowia. W sferze tego typu narzędzi w ostatnich latach doszło do gwałtownego przyspieszenia postępu po długim okresie stagnacji [5]. Maszyny, którymi dysponujemy, zdołały przekroczyć swoje wcześniejsze, wąskie granice i obecnie wykazują się szerokim zakresem umiejętności w rozpoznawaniu wzorów, zaawansowanej komunikacji oraz w innych działaniach kiedyś uznawanych za wyłączną domenę człowieka [5].

W zakresie zastosowania tego typu programów do podejmowania działań w badaniach i decyzjach dotyczących ochrony zdrowia możemy prześledzić dyskusję w perspektywie co najmniej ostatnich 35 lat [6]. Dopiero ostatnio jednak wdrożenie takich rozwiązań stało się faktycznie możliwe. Jest to skutek zarówno samego gwałtownego rozwoju programów posiadających cechy sztucznej inteligencji, jak i generowania przez system opieki zdrowia coraz większej liczby danych, których analiza przekracza zdolności poznawcze człowieka. Pojawienie się nowych rozwiązań technicznych pozwala uzyskać rozległy zakres informacji powiązanych ze stanem zdrowia, a następnie powierzyć je do wspólnego repozytorium, gdzie są dostępne dla szerokiego kręgu odbiorców [6]. Takie repozytoria obejmują diagnozy, charakterystykę terapii i stosowanych leków, wyniki testów laboratoryjnych czy obrazy uzyskane w wyniku specjalistycznych badań, jak na przykład radiologicznych [7]. Łączone mogą być nadto z informacjami o charakterze niezwiązanym bezpośrednio z udzielanymi świadczeniami medycznymi, ale takimi, które mogą wywierać wpływ na ocenę stanu zdrowia, jak na przykład z wynikami sekwencjonowania genetycznego określającymi dziedziczne predyspozycje osoby do wystąpienia określonych cech, charakterystyką ekspresji genów [8], po analizę danych wskazujących na aktywności czy nawet dokonywane zakupy, pozwalające uzyskać informacje o diecie czy trybie życia [4]. Tym samym umożliwiają analizę znacznie szerszego zakresu danych i bardziej szczegółowe dostosowanie świadczenia.

Charakterystyka programów określanych jako sztuczne inteligencje

Pojęcie sztucznych inteligencji w znaczeniu powszechnie wykorzystywanym jest trudne do jednoznacznego i precyzyjnego zdefiniowania na gruncie prawnym. Powyższego terminu nie uwzględnia między innymi Norma ISO 2382-1, która opisuje pojęcia wykorzystywane w technologiach informacyjnych. Pojęcie sztucznych inteligencji jest używane raczej w kontekście opisowym, odnosząc się do „myślących narzędzi" – standardowo podawanym przykładem tego typu aplikacji są tutaj na przykład te wykorzystywane do automatycznego sterowania samochodami [9].

Wspólnym elementem tego typu aplikacji jest ich zdolność do samouczenia – rozumianego jako proces, w którym system automatyczny zdobywa wiedzę, umiejętności bądź dokonuje reorganizacji posiadanych zasobów informacyjnych, mogących być wykorzystanymi do zwiększenia jego efektywności [10]. Używane są do tego „treningowe" próbki informacji, które pozwalają programowi rozpocząć działanie w danym obszarze. Następnie program może zastosować do dalszej nauki kolejne analizowane przypadki, tym samym nieustannie poszerzając posiadaną i wykorzystywaną do podejmowania decyzji bazę danych. Na takiej podstawie program może tworzyć następnie prognozy dotyczące możliwych dalszych elementów – poprzez analogie do posiadanych już kategorii – i rozwoju analizowanej sytuacji. Cechą predykcji dokonywanych przez uczące się programy jest ich probabilistyczny charakter – oparte są, co do zasady, na wyprowadzeniu pewnych prawidłowości istniejących w znanych programowi kategoriach, które następnie odnosi się do nowej sytuacji. Analizy takie mogą przybrać bardzo skomplikowany, wielopoziomowy charakter, zazwyczaj w tego typu programach jednak obecny jest element obejmujący jedynie określenie wysokiego prawdopodobieństwa wystąpienia w danej sytuacji konkretnego zjawiska, nie zaś pewność danego skutku.

Zróżnicowane są natomiast charakterystyki techniczne takich programów, to jest sposób, w jaki wykonują one swoje zadania. Klasycznymi przykładami są na przykład drzewa decyzyjne, w tym o charakterze wspomagającym (*Booster Decision Trees*) i dodatkowym (*Addictive Decision Trees*), sieci interferencji oparte na teorii prawdopodobieństwa subiektywnego (*Bayesian Networks*), programy o charakterystyce ewolucyjnej i sieci neuronowe (*Deep learning neural networks*) [11, 12]. Wskazując na wątpliwości związane z wykorzystaniem programów, autor będzie się odwoływać do ostatniej pośród wyżej wymienionych kategorii, to jest sieci neuronowych. Stwarzają one większe trudności w zakresie uzyskania informacji dotyczących zasad rozumowania programu czy przewidzenia jego rozwoju niż klasyczne drzewa decyzyjne [13, 14]. Taka ich niższa przejrzystość sprawia, że wyraźniej można na ich przykładzie przedstawić niektóre wątpliwości i zagrożenia związane ze stosowaniem programów szeroko rozumianych jako sztuczne inteligencje.

Sieci neuronowe modelowane na wzór ludzkiej sieci nerwowej[4] zawierają w swojej strukturze sztuczne neurony – perceptrony, które są angażowane w sytuacji, gdy ćwiczenie treningowe da wynik pozytywny według określonych wcześniej parametrów, oraz pozostają bierne, kiedy wynik nie spełnia takiego kryterium. Jest to tak zwany trening z nauczycielem (*supervised learning*), który może być prowadzony wówczas, gdy znane są pożądane odpowiedzi sieci na dane sygnały wejściowe. Kolejne dostrajanie parametrów działania neuronu i następne analizy wyników dokonane przez kolejne perceptory ostatecznie prowadzą do uzyskania pożądanych wyników końcowych [15]. Programy tego typu często są wykorzystywane właśnie przy analizie badań, których wyniki można przedstawić w formie obrazu, jak na przykład tomografia komputerowa – gdzie dalsze poziomy aktywacji sztucznych neuronów będą odpowia-

dać rozpoznaniu bądź brakowi rozpoznania obecności kolejnych wyszukiwanych elementów obrazu, następne rzędy ukrytych perceptorów zaś uaktywniane będą w zależności od uzyskanych wartości poprzednich rzędów. Dopiero ostatnia klasa perceptorów to wynik widoczny dla korzystającego z takiego programu.

Wyjątkową cechą programów określanych jako sztuczne inteligencje będzie wysoki poziom ich autonomii, rozumianej jako zdolność do samodzielnego podejmowania decyzji dotyczących zmian wewnętrznych, to jest nie jedynie w wyniku i w zakresie wyznaczonym przez bezpośrednie polecenie użytkownika [16]. W szczególności informacje mające znaczenie dla rozwiązania konkretnego zagadnienia mogą być pozyskiwane przez program samodzielnie z wyznaczonego otoczenia (takiego jak bazy danych poprzednich operacji), bez bezpośredniego wkładu człowieka [17]. Tak na przykład w podanej powyżej sytuacji samouczącej się sieci neuronowej program sam decyduje o kolejnych kategoriach badania, przyporządkowaniu obrazu do określonych synaps oraz o dodawaniu dalszych kategorii synaps dodających późniejsze zaobserwowane zależności, które będą wykorzystywane w następnych, analogicznych badaniach. Nowe dane pobrane w toku nauki mogą prowadzić do wytworzenia nowych klasyfikacji, to jest nowych perceptorów wyższego rzędu, które następnie mogą być przez program wykorzystane do tworzenia klasyfikacji innych niż te, które były we wcześniejszych postaciach programu, te z kolei będą wykorzystywane do opisywania następnych wprowadzanych danych [18]. To właśnie taka niewymagająca ingerencji człowieka możliwość autonomicznego definiowania bądź modyfikowania zasad podejmowania decyzji będzie wyróżniała samouczące się programy.

Samouczące się programy w perspektywie prawa medycznego

Z prawnego punktu widzenia wykorzystanie możliwie najbardziej zaawansowanych form spersonalizowanej medycyny do skuteczniejszego dostosowania terapii do danej osoby jest nie tylko uzasadnione, lecz może być uznane za obowiązek związany z udzielaniem świadczeń medycznych. Jak wskazuje art. 6 ust. 1 Ustawy o prawach pacjenta i rzeczniku praw pacjenta [19], pacjentowi przysługuje prawo do świadczeń zdrowotnych odpowiadających wymaganiom odpowiedniej wiedzy medycznej. Analogiczny wymóg co do standardu udzielanych świadczeń zdrowotnych ujęty jest także w przepisach dotyczących sposobu wykonywania kolejnych zawodów medycznych, jak np. art. 4 Ustawy o zawodzie lekarza i lekarza dentysty (uzl), który mówi o obowiązku wykonywania zawodu zgodnie ze wskazaniami wiedzy lekarskiej i dostępnymi lekarzowi metodami i środkami zapobiegania [20]. Podobny wymóg znajdziemy także na przykład w Kodeksie etyki lekarskiej (kel), który dodatkowo wskazuje w art. 6 zd. 2, iż lekarz powinien ograniczyć czynności medyczne do tych, rzeczywiście potrzebnych choremu zgodnie z aktualnym stanem wiedzy. Wprawdzie nie ma stałych i niezmiennych reguł postępowania medycznego [21], ale o ile za zasadne należy przyjąć uznanie równości wszelkich reprezentowanych kierunków [21], to uwzględniony powinien być indywidualny charakter każdego przypadku [22]. Jeśli brak jest konsensusu co do zakresu wiedzy medycznej, to w celu ustalenia zgodności świadczenia z aktualną wiedzą medyczną powinien być zbadany i wzięty pod uwagę jak najszerszy zakres czynników czy elementów. Analizując wzorzec postępowania lekarza na podstawie art. 4 uzl, orzecznictwo wskazuje, że postępowanie takie powinno przy zachowaniu aktualnego stanu wiedzy gwarantować przewidziany rezultat [23], uwzględnić należy także dane, którymi lekarz dysponuje bądź mógł dysponować [24]. Zasadne wydaje się zatem stwierdzenie, że w celu udzielania świadczeń odpowiadających wymogom współczesnej wiedzy medycznej należy korzystać z możliwie szerokiego jej repozytorium. Na tej podstawie łatwo uznać za uzasadnione stosowanie rozwiązań ułatwiających korzystanie ze wskazań aktualnej wiedzy lekarskiej – w tym programów decyzyjnych.

Samo wykorzystanie programów może jednak budzić poważne wątpliwości o charakterze prawnym i regulacyjnym. Obecnie brak jest jednoznacznych regulacji, które odnosiłyby się do zasad ich wykorzystania. Zgodnie z treścią art. 2 ust. 1 pkt 38 Ustawy o wyrobach medycznych (uwm) [25] są one objęte jej regulacją, o ile przeznaczone są do używania w celach diagnostycznych, terapeutycznych, rehabilitacyjnych i profilaktycznych, wypełniając przesłanki bycia uznanym za aktywny wyrób medyczny. Ustawa ta nie przewiduje jednak rozwiązań pozwalających na kontrolę działania takiego programu czy wymogów dotyczących automatycznie wprowadzanych w nim zmian.

Z pkt 12.1.1 z Załącznika nr 1 do Rozporządzenia Ministra Zdrowia w sprawie wymagań zasadniczych oraz procedur oceny zgodności wyrobów medycznych [26] wynika, iż oprogramowanie takie powinno być walidowane zgodnie z aktualnym stanem wiedzy oraz z uwzględnieniem zasad cyklu rozwoju zarządzania ryzykiem, walidacji i weryfikacji takiego oprogramowania. Cykl ten jednak, zwłaszcza w odniesieniu do weryfikacji, w zakresie, w jakim jest przewidziany w uwm oraz w wyżej wymienionym rozporządzeniu, nie uwzględnia charakterystyki programów o zmiennej strukturze, przewidując jedynie w pkt 3.4 rozporządzenia procedurę dotyczącą zmian planowanych przez wytwórcę programu. Zatem o ile możliwość dokonywania zmian wbudowana jest w samą konstrukcję takiego wyrobu, o tyle pozostaje ona wciąż nieuregulowana.

Wątpliwości związane z działaniem programu

Regulacja samouczących się programów jest niezbędna ze względu na potencjalną niedoskonałość ich zastosowania. Program będzie sam podejmował działania wpływające na jego strukturę, niezależnie od wkładu człowieka. Skutkiem takiego działania może być generowanie bardzo skomplikowanych i niejasnych struktur, a ostatecznie tak zwanych czarnych skrzynek (*black box*) – programów, do których użytkownik, a w niektórych sy-

tuacjach architekt, nie ma dostępu i nie jest w stanie poznać zasad ich działania. Tym samym powstaje ryzyko, że proces decyzyjny takiego programu nie będzie mógł być skutecznie poznany, zbadany oraz zweryfikowany, ani przez osoby, których dotyczył proces decyzyjny, ani przez korzystających zeń [4].

Etyczne wątpliwości dotyczące zasad działania programów przybierają jedną z trzech podstawowych postaci:

- wątpliwości dotyczące procesu zbierania danych;
- wątpliwości związane z samym działaniem podjętym w związku z określonymi danymi;
- wątpliwości związane z odpowiedzialnością za działanie programu [27].

Wątpliwości dotyczące procesu zbierania danych

Niepewności dotyczące zasad zbierania i wyboru danych przez program

Pierwszą z wątpliwości dotyczących zasad stosowania programów samouczących się jest niepewność wskazywanych przez nie wyników. Jak powiedziano wcześniej, ich decyzje i analizy mają charakter probabilistyczny – w trakcie nauki korzystają one faktycznie z danych statystycznych [27]. Jednym ze skutków takiego sposobu nauki będzie poleganie przede wszystkim na współwystępowaniu cech – program nie odróżnia korelacji od związku przyczynowo-skutkowego i opiera się przede wszystkim na wspólnym występowaniu w badanych grupach czynników będących przyczyną i skutkiem. Rezultatem takiego działania mogą być z kolei sytuacje, w których system podając wyniki, działa jedynie na podstawie korelacji pomiędzy zjawiskami, nie analizując istnienia **związku** przyczynowo-skutkowego. Problem ten w szczególnym stopniu może dotyczyć programów o charakterze predykcyjnym [28], które z zasady w większym stopniu korzystają w procesie nauki z korelacji danych w celu wykazania możliwych konsekwencji danej sytuacji, jako że mają ograniczoną możliwość dokonania analizy samej grupy końcowej. Co więcej, takie przypadkowe korelacje mogą dotyczyć całych grup bądź populacji i nie przekładać się bezpośrednio na wartości dotyczące konkretnych osób [27]. Zatem osoba podejmująca pracę z takim programem może otrzymać decyzję, która jest faktycznie oparta na mniej lub bardziej przypadkowych zależnościach pomiędzy danymi.

Niepewność dotycząca weryfikacji danych

Drugim, pokrewnym problemem jest trudność związana z weryfikacją badań, prowadząca do ich niejasności i nieprzejrzystości samego procesu. Niejasność taką można rozumieć przez sytuacje, w których osoba korzystająca z wyników programu nie posiada praktycznej możliwości ustalenia, w jaki sposób program dokonał odpowiedniej klasyfikacji i podjął daną decyzję [27]. Dotyczyć to może nie tylko samego braku możliwości interpretacji, lecz także sytuacji, w której interpretacja taka byłaby nadmiernie utrudniona przez liczbę elementów, które musiałaby dotknąć taka analiza. Aby mówić o możliwości weryfikacji, informacja powinna być dla użytkownika danego systemu *dostępna* oraz *zrozumiała* [27]. Użytkownik musi być w stanie najpierw uzyskać informację, a następnie dokonać analizy i oceny wpływu różnych czynników początkowych na ostateczną decyzję systemu [29]. W obydwu tych aspektach, tak dostępności, jak i zrozumiałości, w przypadku samouczących się systemów czynienia pojawia się pytanie o faktyczną efektywnością sprawowanego nadzoru. Nawet jeżeli dane dotyczące procesu decyzyjnego są faktycznie dostępne, to automatycznie ucząca się maszyna z zasady może posiadać przewagę informacyjną nad jej użytkownikiem – to jest znajdować się w sytuacji, w której praktycznie charakter podejmowanych przez nią działań nie będzie mógł być na bieżąco zweryfikowany przez użytkownika ze względu na szybkość i liczbę prowadzonych na danych operacji [30]. W oczywisty sposób taka niemożliwość czy nawet niezwykła uciążliwość dokonania weryfikacji ustaleń programu może prowadzić do utrudnienia, a w skrajnych przypadkach faktycznego zaniku nadzoru ze strony użytkownika nad jego działaniem – także w takich sytuacjach, w których nadzór nad działaniem jest sprawowany, ale użytkownik nie jest w stanie praktycznie analizować i kontrolować wpływu różnych elementów na ostateczne wyniki działania programu. Problem ten w większym stopniu dotyczy właśnie, jak wskazano uprzednio, sieci neuronowych, jako mniej przejrzystych ze względu na swoją architekturę. Jednak trudności w ocenie zebranych danych ze względu na ich zbyt dużą liczbę mogłyby się pojawić w każdej formie samouczącego się programu.

Oprócz trudności związanych z oceną samej konkretnej decyzji faktycznie taka niejasność i brak możliwości zbadania działania programu mogą uniemożliwić właściwą ocenę ryzyka związanego ze stosowaniem danego programu i związaną z tym zagadnieniem procedurę *oceny ryzyka* o kluczowym znaczeniu dla dopuszczenia go do użytku medycznego i klasyfikacji wyrobu medycznego [31] oraz dla przedstawienia pacjentowi potencjalnych ryzyk związanych z leczeniem, jeżeli zostało ono wyznaczone przez taki program. Zasady działania programu samodzielnie się uczącego powinny być weryfikowane zarówno na etapie wstępnym, to jest samego przyjęcia go do wykorzystania, jak i w przypadku jego ciągłego wykorzystywania, system zaś powinien zachować przez cały okres swojego działania zdolność do takiej weryfikacji.

Niewłaściwe przyporządkowanie kategorii prowadzące do stronniczości

Trzecim problemem dotyczącym zbierania danych jest możliwość wystąpienia stronniczości po stronie takich programów w zakresie zbierania danych, to jest niesłuszne uwzględnienie przy podejmowaniu decyzji określonej cechy w stopniu większym, niż jest to uzasadnione jej faktycznym znaczeniem dla wystąpienia danego skutku. Problem ten występuje wbrew często wskazywanej [32] jako ich zaleta bezstronności procesów decyzyjnych programów komputerowych, które miałyby być wolne od ludzkich uprzedzeń i podejmowane w sposób całkowicie

obiektywny [32]. Niewątpliwie programy mogą zachować wartości i uprzedzenia architekta systemu, następnie powielane przez samouczący się algorytm w kolejnych elementach – poprzez niewłaściwą ocenę wagi przykładów i odpowiednie, niewłaściwe uwzględnienie ich na dalszych etapach nauki [32]. Takie uprzedzenia mogą być wprowadzane w sposób bezpośredni w przypadku zakodowania w programie niewłaściwych zasad podejmowania decyzji. Mogą także zostać wytworzone przez sam program, na przykład poprzez dokonanie analizy danych niereprezentatywnych i mających silne tendencje ku określonej charakterystyce, co prowadzi do nieuniknionego odzwierciedlania przez program uprzedzeń [33]. Uprzedzenia mogą się ujawnić także w związku z określonymi oczekiwaniami społecznymi, wpływającymi na analizowane przez program a także ze względu na różnice w charakterze i zakresie informacji dotyczących osób analizowanych przez program w różnych przypadkach. Dalsze uprzedzenia mogą wynikać z postępu w nauce, na przykład programy diagnostyczne typu CDSS[5] często wykazują uprzedzenie wobec metod leczenia zawartych od początku w ich architekturze w stosunku do nowych metod wskazywanych później [34] – ze względu na zapisanie wcześniejszych metod w większej liczbie utworzonych procesów decyzyjnych i odgrywanie przez nie tym samym większej roli w rozumowaniu takiego programu.

Wątpliwości dotyczące działań podejmowanych przez program

Dyskryminacja

Zagrożenie dyskryminacją w wyniku podejmowania przez zautomatyzowany algorytm decyzji może być skutkiem określonego modelu podejmowania decyzji. W zakresie, w jakim dotyczy ono wskazywania na pewne zależności w związku z zebranymi danymi, polega na zbieraniu informacji i dopasowywaniu osób do odpowiadających im profili osób, które program taki wytworzy [35]. W tej sytuacji dalsze decyzje dotyczące konkretnej osoby mogą być podejmowane przez program na podstawie cech charakteryzujących statystycznie znaczącą część danej grupy tworzącej profil, nawet jeżeli nie znajduje ona bezpośredniego zastosowania do samej osoby, której decyzja dotyczy [36]. Główną różnicą między zagrożeniem o charakterze epistemicznym, dotyczącym stronniczości zbierania danych, a dyskryminacją będzie to, że dyskryminacja odnosi się tutaj do długofalowych skutków takiego podejmowania decyzji przez programy, prowadząc do sytuacji, w których wskutek powtarzalnych określonych kwalifikacji osób niektóre tylko grupy będą faktycznie posiadały dostęp do konkretnych informacji bądź możliwości, wzmacniając tym samym istniejące już podziały społeczne [27]. W miarę kolejnych wykorzystań takiego programu podziały te mogą się nasilać, wskutek powtarzalności informacji wstępnie skutkującej zagrożeniem dyskryminacją. Początkowe decyzje, wpływałyby na okoliczności, których dotyczy działanie programu, prowadząc do powtarzania pierwotnego wyniku ze względu na korzystanie ze stronniczej bazy danych.

Dotyczyć to może między innymi takich sytuacji, w których w stosunku do określonych grup występują nierówności informacyjne. Rozumieć przez to należy sytuację, w której już istniejące nierówności i różnice społeczne wpływają na zakres zbieranych informacji dotyczących grup. Takie różnice w zakresie zebranych informacji mogą następnie wypaczyć działanie programu w odniesieniu do określonych grup[6]. Z odmienną formą takiego zagrożenia dyskryminacją możemy mieć do czynienia wówczas, gdy przynależność do określonej grupy będzie powiązana z potencjałem wystąpienia określonego skutku bądź negatywnego wyniku[7].

Program dokonujący zbierania danych może być zabezpieczony przed utworzeniem dyskryminacyjnego charakteru takich działań poprzez zastosowanie czterech strategii [39]. Pierwszą z nich będzie utrzymywanie kontroli nad danymi wykorzystywanymi przez program do nauki na etapie poprzedzającym samo jego wykorzystanie, przykładowo obejmujące zarówno pomijanie niektórych elementów, kontrolę nad elementami, które mogą być uznane za prowadzące do dyskryminacji i ich odpowiednie ważenie czy wykorzystywanie dodatkowych kategorii danych zbieranych, które zmniejszą szanse wystąpienia dyskryminacji [39]. Drugą formą takiego działania może być integracja kryteriów antydyskryminacyjnych do rozumowania algorytmu poprzez uzupełnienie go o dalsze działania czy kryteria antydyskryminacyjne pozwalające na usunięcie potencjalnych skutków zbierania danych [39]. Trzecią formą będzie wdrożenie kolejnych procesów przetwarzających dane w taki sposób, aby usunąć potencjalne dyskryminacyjne skutki, już po uzyskaniu wyników. Czwartą formą działań antydyskryminacyjnych mogłoby być wprowadzenie, bez ingerencji w samo działanie programu, rozwiązań zapewniających proporcjonalność decyzji w taki sposób, by w przypadkach granicznych w grupie dotkniętej dyskryminacją dokonać weryfikacji podjętych decyzji na pozytywne [40].

Wątpliwości związane z poszanowaniem autonomii jednostek

Wykorzystywanie zautomatyzowanego przetwarzania danych prowadzi także do wytworzenia poważnych wyzwań dla poszanowania autonomii i praw osób, których dane dotyczą.

Zautomatyzowane programy dokonujące analiz danych osoby z zasady podejmują decyzje poprzez zaliczenie takiej osoby do określonych grup, tworzonych według określonych cech i profili takiej osoby [27]. Tym samym działania programów zautomatyzowanych są z prawnego punktu widzenia klasyfikowane jako tak zwane profilowanie osoby fizycznej. Takie działanie zdefiniowane może być jako forma przetwarzania danych dotyczących osoby, która polega na wykorzystaniu dostępnych informacji dotyczących osoby do analizowania i przewidywania aspektów tej osoby, czyli na wyprowadzaniu wniosków jej dotyczących [41]. Takie profilowanie może, co do zasady, przybrać jedną z dwóch form:

- analizowania i przewidywania aspektów dotyczących osoby fizycznej poprzez zestawianie z sobą informa-

cji dotyczących tej osoby (można je określić jako opracowanie profilu);
- przypisywanie osobom dodatkowych cech, pochodzących z analizy statystycznej informacji dotyczących innych osób (można je określić jako przypisanie profilu) [41].

Zagrożeniem bezpośrednio powiązanym z profilowaniem jest pewne rozmycie się indywidualnych cech jednostki na etapie przyporządkowania takiej jednostki do określonej grupy [27]. Automatycznie podejmujące decyzje programy odrzucają pewien zakres cech jednostki, które uznawane są za pozbawione znaczenia dla całokształtu sytuacji, tym samym niejako redukują ją jedynie do cech związanych z jej profilem. Oddzielną wątpliwością dotyczącą wpływu na autonomię osoby, której dotyczy badanie, może być poszanowanie konkretnych praw pacjenta w przypadku wykorzystywania samouczących się programów. Przykładem takich właśnie trudności związanych z realizacją praw jednostki w przypadku programów nieprzejrzystych może być kwestia wykonania prawa pacjenta do zasięgnięcia opinii innego lekarza bądź zwołania konsylium z art. 6 ust. 3 Ustawy o prawach pacjenta i Rzeczniku Praw Pacjenta (upp) w przypadku opinii wydanej na podstawie diagnozy dokonanej przez taką automatyczną medyczną sztuczną inteligencję. Jeżeli opiniowanie w konsylium bądź przez innego lekarza odbywa się z wykorzystaniem programów analitycznych, trudno mówić o ponownym rozpatrzeniu sytuacji pacjenta, jeżeli ponowne diagnozowanie odbędzie się z wykorzystaniem tego samego programu. Celem tego uprawnienia jest weryfikacja początkowo wydanej diagnozy [21]. Osiągnięcie takiego celu byłoby niemożliwe w sytuacji, w której kolejny lekarz bądź też konsylium wydający decyzję zmuszeni są oprzeć się na wynikach dostarczonych przez ten sam program. Zakres, w jakim delegacja z art. 6 ust. 3 upp daje pacjentowi uprawnienie do wskazania określonej osoby, której opinii należałoby zasięgnąć [21], w wyraźny sposób wskazuje na interes pacjenta obejmujący element zaufania do źródła opinii. W tej sytuacji w celu zapewnienia prawa do zasięgnięcia dodatkowej opinii medycznej w przypadku wykorzystania automatycznego programu diagnostycznego niezbędne będzie przeanalizowanie rozumowania takiego algorytmu i zaopiniowanie dokonanej klasyfikacji przez osoby ponownie podejmujące decyzję. Jeżeli natomiast podjęlibyśmy próbę całkowitego usunięcia elementu zautomatyzowanego z podejmowania drugiej decyzji, mogłaby ona być w znaczącym stopniu odmienna, ze względu na o wiele bardziej ograniczony zakres danych, na podstawie których została podjęta.

Wątpliwości związane z odpowiedzialnością za działania programu

Ostatni problem związany ze stosowaniem zautomatyzowanych programów podejmujących decyzje obejmuje samą kwestię przyporządkowania odpowiedzialności za jakiekolwiek błędne bądź wadliwe decyzje podejmowane przez program. W przypadku samouczących się programów trudno wskazać, jaka konkretnie osoba jest odpowiedzialna za powstałe błędy. Nieskuteczne są tutaj standardowe zasady przypisania odpowiedzialności. Analizując, przykładowo, program o charakterze diagnostycznym, wskazać należy, że w systemie ochrony zdrowia osobą odpowiedzialną za błędną diagnozę będzie, co do zasady, przedstawiciel zawodu medycznego udzielający świadczenia. To właśnie ta osoba przejmuje obowiązki dotyczące zasad udzielania świadczenia medycznego, w tym staje się gwarantem w rozumieniu przepisów prawa karnego[8]. Jest też osobą, na którą nałożony został obowiązek podejmowania działań zgodnie ze wskazaniami aktualnej wiedzy medycznej według odpowiednich przepisów ustaw zawodowych[9].

Warto tutaj wskazać także na fakt, że odpowiedzialność gwaranta za zaniechanie nie jest ograniczona wyłącznie do błędu, ale do każdej sytuacji, w której zaniechanie dokonania czynności prowadzi do narażenia pacjenta na bezpośrednie niebezpieczeństwo utraty życia albo ciężkiego uszczerbku na zdrowiu [42]. Na gruncie tej zasady, przykładowo, lekarz mógłby być odpowiedzialny za błąd w sytuacji, w której nie dokonał weryfikacji diagnozy wskazanej przez program. Z kolei analizując samą zasadę odpowiedzialności związanej z wykorzystaniem takiego programu, zarzut dotyczący postępowania i niedokonania analizy można faktycznie postawić jedynie wtedy, gdy dana osoba miała jakąś możliwość kontroli procesu decyzyjnego programu bądź wejrzenia w tę decyzję i jej przeanalizowania [27]. Zgodnie z zasadą art. 30 Kodeksu karnego, aby przypisać winę, niezbędnym elementem jest zdolność osoby do rozpoznania bezprawności danego czynu. Trudno mówić o tym, by lekarzowi korzystającemu z automatycznego programu analitycznego, w sytuacji, w której program taki błąd popełnia, a lekarz nie posiada możliwości dokonania weryfikacji, można było postawić zarzut takiej świadomości co do bezprawności.

Zawodzi tutaj stosowany faktycznie w kontekście odpowiedzialności zawodowej – w tym art. 6 uzl – wzorzec staranności. Nawet postępując zgodnie ze wskazaniami wiedzy medycznej, w myśl art. 6 uzl oraz art. 4 kel, do oceny decyzji zgodnie z najnowszymi wskazaniami wiedzy lekarskiej konieczna będzie możliwość samego wejrzenia w strukturę decyzyjną takiego programu i dokonanie analizy sposobu jej podjęcia; wymagać to może nie tylko wiedzy medycznej, lecz także umiejętności oceny zasad podjęcia decyzji. Czynność taka mogłaby być zarówno niezwykle czasochłonna, jak i wręcz niemożliwa do wykonania; poza tym nie można wskazać, by w zakresie „postępowania zgodnego ze wskazaniami wiedzy medycznej" mieściła się umiejętność oceny zasad działania programu komputerowego, w tym zwłaszcza sposobu dokonywania przez niego klasyfikacji. Zgadza się to ze stanowiskiem Sądu Najwyższego – dla odpowiedzialności lekarza niezbędne jest obiektywne przypisanie mu skutku należącego do znamion strony przedmiotowej, objętego tym przepisem przestępstwa, takiego że gdyby podjął nakazane czynności, to skutek by nie nastąpił [43]. Zatem niezbędna dla takiej odpowiedzialności byłaby sytuacja, w której pominięto jakąś czynność, pozwalającą na uzyskanie oczekiwanego, pozytywnego skutku w postaci

uzyskania wiedzy o danej sytuacji. Nie można mówić o takim braku podjęcia nakazanych czynności wówczas, gdy obiektywnie lekarz nie był w stanie zweryfikować procesu decyzyjnego.

Jednocześnie trudno jest także wskazać na inny podmiot odpowiedzialny na ogólnych zasadach w takim przypadku. Błąd programu nie musi wynikać bezpośrednio z wadliwej jego konstrukcji – uprzedzenia bądź preferencje programu, jak wskazano uprzednio, będą czymś, co będzie się pojawiać w trakcie jego wykorzystywania oraz tworzenia nowych reguł i zasad rozumowania. Ostatecznym rezultatem może być tutaj sytuacja, w której faktycznie nikt nie posiada kontroli nad działaniami maszyny w takim stopniu, by możliwe było przypisanie odpowiedzialności za jej działanie [30]. Faktycznie obecnie brak jest jakichkolwiek rozstrzygnięć, które jednoznacznie określałyby takie zasady odpowiedzialności.

Etyczna architektura programu i prawo do wyjaśnienia

Analizując przedstawione wyżej wątpliwości, należy podkreślić, że wspólnym elementem występującym we wszystkich aspektach epistemicznych oraz normatywnych jest niejednoznaczność i brak przejrzystości komputerowego programu oraz trudność z uzyskaniem informacji dotyczącej tego, w jaki sposób bądź dlaczego program podjął daną decyzję [44], tym samym faktycznie uniemożliwiając jakąkolwiek kontrolę czy weryfikację. Częściowym rozwiązaniem wymienionych wyżej wątpliwości mógłby być wymóg zapewnienia należytej przejrzystości programu, rozumianej jako wprowadzenie właściwych warunków dostępu do informacji dotyczącej jego działania oraz tego, w jaki sposób uzyskana z programu informacja może wpłynąć na podejmowanie decyzji samego programu oraz jego użytkownika [45].

W zakresie architektury programów postulat ten ujęty jest w propozycji stworzenia „Prawa do wyjaśnienia", które obejmowałoby nałożony na producentów obowiązek udostępnienia informacji dotyczącej tego, w jaki sposób one funkcjonują i podejmują decyzje. O ile adresatem tego uprawnienia byłaby osoba, której dane są przetwarzane, o tyle trudno wyobrazić sobie sytuację, w której użytkownik programu nie posiadałby dostępu do takiej informacji, posiada ją zaś osoba, której dane były przez program przetwarzane. Należy także zwrócić uwagę na fakt, że jakiekolwiek tego typu ograniczenie dotyczyłoby jedynie programów przetwarzających dane osobowe, to jest takich, dzięki którym można zidentyfikować osoby, których te informacje dotyczą. „Prawo do wyjaśnienia" nie wpływałoby zatem w żaden sposób na architekturę takich programów, które działają w sposób abstrakcyjny.

Częściowo do postulatu dotyczącego tego prawa odniósł się ustawodawca europejski w przepisach Rozporządzenia Parlamentu Europejskiego i Rady (UE) 2016/679 z 27 kwietnia 2016 roku w sprawie ochrony osób fizycznych w związku z przetwarzaniem ich danych osobowych i w sprawie swobodnego przepływu takich danych (RODO) [46], przyjmując uregulowania ograniczające możliwość tworzenia algorytmów typu „czarnej skrzynki" oraz wskazujące na konieczność udostępnienia i udzielania informacji osobom, których dotyczy działanie takich programów. Analizując przyznane na gruncie RODO uprawnienia do uzyskania wyjaśnienia dotyczącego zasad działania zautomatyzowanego programu podejmującego decyzje, należy się odnieść do możliwych przyczyn niejasności związanej ze strukturą samodzielnie uczących się programów.

Wskazać tutaj można trojakie źródło wpływające na powstawanie przeszkód w nadaniu programowi należytej przejrzystości i udostępnieniu użytkownikowi odpowiednich informacji dotyczących zasad jego funkcjonowania:

- Informacja taka może być celowo zatajona alternatywnie przez producenta bądź też właściciela systemu, który nie ujawnia publicznie i nie przekazuje korzystającym zasad działania takiego systemu.
- W przypadku, w którym nastąpiło udostępnienie takiej informacji, może dojść do wystąpienia luk w wiedzy technicznej po stronie odbiorcy, utrudniających bądź uniemożliwiających zrozumienie informacji dotyczących przetwarzania i zasad działania systemu.
- Poziom skomplikowania zasad działania takich rozwiązań może być nadmierny w stosunku do ludzkich możliwości poznawczych i interpretacyjnych, co uniemożliwiałoby przedstawienie takiej informacji. Taki nadmierny poziom skomplikowania może być zwłaszcza rezultatem zautomatyzowanej optymalizacji wielowymiarowych charakterystyk samouczących się programów komputerowych [44].

Celowe zatajenie informacji dotyczącej działania programu

Pierwsza z wyżej wymienionych przeszkód może wystąpić zwłaszcza w sytuacji, w której architektura programu jest utrzymywana w tajemnicy w celu osiągnięcia przewagi nad konkurencją [44]. Może też wynikać z samej części zatajenia pewnych elementów programu, na przykład faworyzujących określone metody i wyroby dostarczane przez danego producenta. Do tej sytuacji bezpośrednio odnosi się tekst RODO w zakresie, w jakim w art. 13 ust. 2 lit. f oraz analogicznym art. 14 ust. 2 lit. g wskazuje na obowiązek udzielania informacji dotyczącej algorytmów. Każdej osobie fizycznej, której dane są przetwarzane w sposób zautomatyzowany, należy udzielić informacji. Tym samym nie jest dozwolone korzystanie z programów, w przypadku których nie jest możliwe udzielenie takiej informacji. Analizując treść tego przepisu, w celu ustalenia, jaki zakres informacji faktycznie powinien być udostępniony, na wstępie należy wskazać, że z jego treści nie wynika, czy uprawnienie to należy odnieść do „istotnych zasad" podejmowania decyzji na podstawie profilowania, czy też do istotnych zasad samego profilowania, to jest wykonywania samej czynności nawet wówczas, gdy taka decyzja nie jest podjęta [47]. Za tym drugim stanowiskiem przemawia literalna analiza angielskiej treści RODO, która w miejscu polskiego zwrotu istotnych informacji o zasadach podejmowania decyzji posługuje się pojęciem *meaningful information about logic involved*. Natomiast za uznaniem, że chodzi tu o profilowanie związane z podejmowaniem decyzji, przemawia wykładnia celowościowa [47]. Z punktu wi-

dzenia programów medycznych rozróżnienie to ma fundamentalne znaczenie. Jak wskazano powyżej, programy wykorzystywane w kwestiach medycznych działają poprzez profilowanie osób, których dane dotyczą. Natomiast o podejmowaniu decyzji mówimy wówczas, gdy wynikiem operacji jest wywołanie wobec osoby skutków prawnych bądź w odmienny sposób istotne wpłynięcie na taką osobę. W przypadku danych dotyczących stanu zdrowia art. 22 ust. 1 w związku z art. 22 ust. 4 RODO wskazuje, że aby osoba podlegała decyzji automatycznej, powinna być ona podejmowana z udziałem człowieka – zatem dla oceny, czy mamy do czynienia z programami podejmującymi decyzje, nie będzie miał znaczenia sam fakt zaangażowania osoby w proces decyzyjny.

Analizując pojęcie decyzji podejmowanej przez program, Grupa Robocza, organ doradczy Komisji Europejskiej w sprawach związanych z ochroną danych, w art. 29 wskazała, że o takim podejmowaniu decyzji w sposób automatyczny można mówić między innymi w takich sytuacjach, gdy skutki takiego działania nie będą trywialne i w znaczący sposób będą wpływały na okoliczności, zachowanie czy wybory osoby [48]. Zasadą przy działaniu programów medycznych będzie dostarczenie informacji, wykorzystywanej przy podejmowaniu dalszych decyzji dotyczących postępowania z taką osobą. W określonych sytuacjach taka informacja wpłynie w znaczącym stopniu na okoliczności, zachowanie czy wybory osoby – zwłaszcza gdy będzie dotyczyć diagnozy. Z kolei program o charakterze predykcyjnym, wskazujący jedynie na możliwość wystąpienia jakichś skutków bądź kwalifikujący osobę do grupy o określonym ryzyku, będzie taki skutek wywierał w znacząco mniejszym stopniu, a w pewnych sytuacjach, takich jak na przykład analiza wzorów i tendencji rozwoju tkanki mózgowej, w ogóle nie wystąpi poza szczególnymi wypadkami. Należy zatem uznać, że o decyzyjności programów medycznych można mówić jedynie w określonych sytuacjach, ich charakter zaś powinien być analizowany oddzielnie dla każdego przypadku.

Jeżeli chodzi o zakres informacji, które mają być udostępnione jednostce w kontekście uprawnienia z art. 13 ust. 2 pkt f, to – jak już wskazano – powinny one dotyczyć istotnych zasad podejmowania decyzji. Odnosząc to pojęcie do samego programu, należy uznać, że dotyczy ono co najmniej logiki wykorzystywanej do podejmowania decyzji, znaczenia zamierzonych skutków przetwarzania, ogólnych zasad wykorzystania systemu oraz, potencjalnie, na przykład samego drzewa podejmowania decyzji w takim zautomatyzowanym systemie, przyjętych sposobów klasyfikacji osoby i tego, w jaki sposób zaliczono ją do określonych kategorii [49]. Natomiast w przypadku gdy informacja dotyczyłaby podejmowanej decyzji, katalog ten należałoby rozszerzyć o przekazanie informacji, które w danej sytuacji uzasadniały i posłużyły do jej podjęcia, przykładowo, poprzez porównanie charakterystyk, zasad podejmowania decyzji przez urządzenie mające znaczenie dla danej sytuacji czy wskazanie grup profilowych, do których osoba została zaliczona – zatem na przykład w przypadku umieszczenia osoby w grupie wysokiego ryzyka wystąpienia określonego schorzenia wskazanie profili, które o takim wysokim ryzyku świadczą [49]. Działanie takie może być dodatkowo uzupełnione przez informacje kontrfaktyczne – to jest poprzez zaprezentowanie przykładowej sytuacji, w której podjęta zostałaby decyzja o odmiennym charakterze [50]. Zatem w przypadku przekroczenia określonej normy, na przykład poziomu hormonów, należałoby tutaj mówić o podaniu takiej osobie przyjętej dla danej sytuacji normy bądź normy, do której mogłaby się odnieść.

Analizując wymienione wyżej elementy, należałoby przyjąć – w opinii autora – że producent programu, który podjął decyzję o zakwalifikowaniu osoby do grupy pacjentów wysokiego ryzyka wystąpienia określonego schorzenia, powinien umożliwić udostępnienie co najmniej następujących kategorii informacji:

- zakres pozyskanych informacji, na których podstawie podjęto decyzję – zarówno wstępnych, jak i takich, z którymi dany przypadek został połączony;
- znaczenie tych informacji i sposób, w jaki informacje te zostały ocenione, w tym sposób podejmowania decyzji;
- charakterystyka grup profilowych, do których program zaliczył daną osobę, na przykład szanse wystąpienia w dane grupie określonych skutków;
- porównanie grup profilowych z grupą wzorcową – przez wskazanie kontrfaktycznych właściwości wzorca.

Kategorie te gwarantują zachowanie przynajmniej pewnego podstawowego poziomu przejrzystości architektury programu. W przypadku, gdyby zostały ujawnione, możliwe będzie dokonanie potencjalnej weryfikacji decyzji podjętej przez program i odniesienie się do aktualnie przyjmowanej metody jego rozumowania.

Luki w wiedzy po stronie odbiorcy

Analizując drugą z podniesionych przeszkód, należy ją odnieść do postulatu, zgodnie z którym można mówić o transparentności procesu, jeśli dotycząca go informacja udzielana jednostce jest zrozumiała [45]. Przeszkoda ta pojawi się w sytuacji, w której zamiarem producenta jest wprawdzie ujawnienie zasad działania programu, jednakże takie ujawnienie pozostaje nieskuteczne z punktu widzenia odbiorcy. Kwestia ta, w sytuacji, w której ma dojść do ujawnienia informacji za zasadzie art. 13 ust. 1 lit f i art. 14 ust. 2 lit. g RODO, została z kolei uregulowana w art. 12 ust. 1 zdanie pierwsze rozporządzenia. Przepis ten wskazuje, że informacje dotyczące czynności wykonywanych na danych – a zatem także dotyczące zasad działania systemu – powinny być przekazywane w formie zwięzłej, przejrzystej, zrozumiałej i łatwo dostępnej jasnym i prostym językiem. Rozumieć to należy w szczególności jako nakazujące stosowanie pojęć jednoznacznych, to jest niebudzących wątpliwości co do znaczenia i treści informacji [47]. Informacja powinna być także łatwa do zrozumienia, przyswojenia i tak sformułowana, aby nie wymagała specjalistycznej wiedzy po stronie odbiorcy informacji [47].

Jako że brak jest sprecyzowania co do poziomu wiedzy, do którego dostosowane powinny być te informacje, znaczenie będzie miał tutaj krąg jej odbiorców. Zatem w przypadku programu przetwarzającego dane pacjen-

tów sposób jej przekazania powinien być dostosowany do ich zdolności poznawczych. Warto wskazać, że zgodnie z przepisami RODO obowiązek dostarczenia takich informacji spoczywa na administratorze danych, którym w przypadku wykorzystywania danych leczniczych będą osoby wykonujące zawód medyczny, a w zakresie zasad działania powinny być one możliwe do udostępnienia. Zatem jeżeli chodzi o sam sposób przekazania informacji, uzasadnione byłoby tutaj zastosowanie zasad udzielenia informacji wskazanych w art. 9 upp czy też w odpowiednich ustawach dotyczących zawodów medycznych. Osoba wykonująca zawód medyczny mogłaby przekazać informacje w sposób odpowiedni po ocenie możliwości intelektualnych i wiedzy osoby, której ich udziela [51]. Tym samym wytwórca programu mógłby umożliwić jej przekazanie takich informacji właśnie poprzez dostarczenie ich w sposób możliwy do interpretacji przez specjalistę.

Niemożność analizy struktury programu

Trzecia z wymienionych przeszkód w udzieleniu informacji stwarza większe trudności, jako że jej wystąpienie funkcjonalnie uniemożliwi wykonanie obowiązków informacyjnych. Będzie ona miała szczególne znaczenie właśnie w przypadku stosowania niektórych samouczących się programów ze względu na brak wyraźnej interpretowalności, uczące się maszyny stoją tutaj na unikatowej pozycji [52]. Programy takie mogą bowiem zazwyczaj modyfikować zasady swojego zachowania w trakcie wykonywania operacji [53]. Kolejnych trudności może tutaj dostarczyć probabilistyczne łączenie takich zasad rozumowania programu w struktury jeszcze bardziej skomplikowane [54]. Ostatecznie może to prowadzić do przesłonięcia zasad działania algorytmu i powstania programu, w którego przypadku wyjaśnienie zasad działania nie jest możliwe, jako że nawet jego producent nie będzie dysponował odpowiednią wiedzą.

Częściowo do tego zagadnienia będzie się odnosić przyjęta w RODO koncepcja ochrony prywatności w fazie tworzenia projektu, to jest *privacy by design.* Zgodnie z treścią art. 25 ust. 1 RODO już na etapie projektowania danej czynności administrator powinien wdrożyć środki techniczne i organizacyjne, zaprojektowane w celu skutecznej realizacji zasad ochrony danych osobowych. Jednocześnie sam system przetwarzania danych powinien być zaprojektowany w taki sposób, by umożliwić efektywną realizację takich właśnie zasad ochrony danych. Tym samym zagadnienia i aspekty związane z ochroną prywatności powinny zostać włączone między innymi w projektowanie procesu oraz w zarządzanie technologiami informacyjnymi i systemami przez cały cykl życia informacji [55], ochrona prywatności zaś powinna stać się częścią składową takiego systemu [47].

Zasada ta faktycznie zakazywałaby utworzenia takiego programu, którego sama architektura uniemożliwia wykonanie obowiązków informacyjnych z art. 13 i art. 14 w sposób jasny i przejrzysty. Zatem w zakresie, w jakim producent programu zamierza podjąć prace projektowe, powinien wprowadzić środki, które wyłączą możliwość powstania „czarnej skrzynki". Niezgodność z przepisami dotyczącymi prywatności nie może być uzasadniona względami natury organizacyjno-finansowej[10] – zatem sama korzyść ze stosowania określonego rozwiązania nie może uzasadniać naruszenia zasad ochrony prywatności. Alternatywą wobec takiego upraszczania konstrukcji algorytmów, które mogłoby być zagrożeniem dla ich rozwoju i funkcjonalności [9], byłaby możliwość wykorzystania przez twórcę danego programu tak zwanych strażniczych sztucznych inteligencji (*Guardian AI*), które stanowiłyby drugą, w stosunku do „normalnych", podejmujących decyzje, kategorię programów komputerowych. Ich rolą byłoby badanie, kontrola, audytowanie oraz nadzorowanie zgodności funkcjonowania operacyjnych sztucznych inteligencji [9], prowadzone poprzez analizę ich zasad działania. Mogłyby one dostarczyć zarówno informacji dotyczącej samych zasad działania operacyjnej sztucznej inteligencji, jak i ewentualnych, dokonywanych przez nią decyzji. Rozwój takiego programu w stosunku do normalnych sztucznych inteligencji mógłby być ograniczony – ze względu na ograniczony zakres przedmiotu badań nie wymagałby on aż tak skomplikowanej struktury jak pierwotna aplikacja analityczna. Warto jednak wskazać, że i taki program sam byłby narażony na wskazane wcześniej zagrożenia dotyczące niejasności zasad zbierania danych.

Podsumowanie

Wzrastająca liczba danych generowanych przez system ochrony zdrowia wymusza, w celu ich efektywnego wykorzystania, stosowanie rozwiązań umożliwiających szybką i skuteczną ich analizę. Od przyjętych regulacji będzie zależeć przeważający model stosowania programów służących do takiej analizy w wielu kwestiach medycznych. Ostatecznym celem ewentualnej regulacji powinno być osiągnięcie pewnej równowagi między korzyściami wynikającymi ze stosowania tego typu programów a związanymi z nimi zagrożeniami. Na jednym końcu spektrum potencjalnych rozwiązań normatywnych możemy znaleźć klasyczny model łatwo zrozumiałego, łatwo potwierdzonego, szeroko wspartego dowodami związku biologicznego, przejrzystego dla lekarzy, naukowców oraz – w idealnym przypadku – dla pacjentów. Na drugim końcu znajdziemy całkowicie nieprzejrzysty model decyzyjny, wyprowadzony z nietransparentnego procesu, będący w stanie tworzyć predykcje, ale niemożliwy do zrozumienia. Pośrodku znajduje się jakaś forma łączona, w której mamy do czynienia z programem bądź całkowicie, bądź częściowo przejrzystym, w którego przypadku możliwe są do poznania niektóre jego aspekty lub funkcje [4].

Przypisy

[1] Przykładem rozpoznania kwestii personalizacji medycyny może być np. uwzględnienie jej w przemówieniu prezydenta Obamy o Stanie Unii z roku 2015 [1].

[2] Prostym przykładem znaczenia personalizacji leczenia może być zasadność stosowania antykoagulantu warfaryny,

w którego przypadku występują znaczące różnice w szybkości metabolizmu u różnych osób. Badania genetyczne prowadzone w ostatnich latach pozwoliły ustalić, iż największe znaczenie dla tego procesu ma przede wszystkim charakter wariacji dwóch genów: cytochromu *P450 2C9 CYP2C9* oraz reduktazy epoksydowej witaminy K VKORC1. Tym samym zasadność podania może być ustalona dla konkretnej osoby w odniesieniu do jej struktury genetycznej w miejsce poprzedniej obserwacji [2].

[3] Sytuacja taka ma mniejsze znaczenie w Polsce. Dla porównania badanie przeprowadzone w roku 2011 przez amerykańską FDA (*Food and Drug Administration*) wykazało, że średnio ponad 90% osób biorących udział w badaniach skutków leków identyfikowało się jako „biali", co mogło znacząco utrudniać zaobserwowanie potencjalnego wpływu leków na niektóre społeczności [3].

[4] Początkowo sieci neuronowe były zaprezentowane jako programy służące do symulacji działania mózgu. Dopiero później zaczęły być wykorzystywane do tworzenia nielinearnych modeli statystycznych [14].

[5] *Clinical Decision Support System* – programy diagnostyczne, łączące obserwacje dotyczące pacjenta z informacjami dotyczącymi danej kategorii pacjentów.

[6] Problemem może być tutaj sytuacja różnicy w dostępności informacji. Dyskryminacja może przykładowo pojawić się w przypadkach, gdy program nie może brać pod uwagę czynników, które nie są rejestrowane wskutek szczególnej sytuacji socjoekonomicznej osób, pozbawiając je możliwości partycypacji w operacjach związanych ze zbieraniem danych czy wypaczających je [37].

[7] Przykładem, o charakterze historycznym, może być tutaj istniejąca w latach 90. XX wieku tendencja do genetycznej predyspozycji do niedoboru żelaza związana z występowaniem niektórych chorób o podłożu genetycznym [38].

[8] Zgodnie z obowiązkiem lekarza określonym w art. 30 uzl: Lekarz ma obowiązek udzielać pomocy lekarskiej w każdym przypadku, gdy zwłoka w jej udzieleniu mogłaby spowodować niebezpieczeństwo utraty życia, ciężkiego uszkodzenia ciała lub ciężkiego rozstroju zdrowia, oraz w innych przypadkach niecierpiących zwłoki.

[9] Art 4 uzl; analogiczne zapisy zawierają art 11 Ustawy z dnia 15 lipca 2011 r. o zawodach pielęgniarki i położnej, art. 21 Ustawy z dnia 27 lipca 2001 r. o diagnostyce laboratoryjnej, art. 4 Ustawy z dnia 25 września 2015 r. o zawodzie fizjoterapeuty.

[10] Zob. wyrok NSA z dnia 4 marca 2002 r. sygn. At II SA 3144/0: żadne względy natury organizacyjno-finansowej nie powinny być traktowane jako podstawy do sprzecznego z prawem przetwarzania danych osobowych.

Piśmiennictwo

1. Obama B., *Sixth Presidential State of the Union Address,* Washington 2015, http://www.americanrhetoric.com/speeches/stateoftheunion2015.htm (dostęp: 20.01.2015).
2. *Warfarindosing*, http://www.warfarindosing.com (dostęp: 1.12.2017).
3. FDA (Food and Drug Administration)*, Collection, Analysis and Avaliability of Demographic Subgroup Data for FDA Approved Medical Products*, Maryland 2013, https://www.fda.gov/downloads/RegulatoryInformation/LawsEnforcedbyFDA/SignificantAmendmentstotheFDCAct/FDASIA/UCM365544.pdf (dostęp: 6.01.2018).
4. Nicholson W., Price II, *Black box medicine*, „Harvard Journal of Law & Technology" 2015; 28 (2): 419–467.
5. Brynjolfsson E., McAffee A., *Drugi wiek maszyn*, (ang. *The second machine age*), tłum. B. Sałbut, MT Biznes, Warszawa 2013.
6. Chandrasekaran B., Mittel S., *Conceptual representation of medical diagnosis for computers. MDX and related systems*, w: *Advances in Computers*, New York Place: 217–293.
7. Jensen P.B, *Mining electronic health records: Towards better research applications and clinical health*, „Nature Review Genetics" 2012; 13: 395–405.
8. Burke W., Bruce M.O.P., *Personalized medicine in era of Genomics*, „Journal of American Medical Association" 2007; 14: 1682–1684.
9. Etzioni A., Etzioni O., *Keeping AI legal, Vanderbilt*, „Journal of Entertainment & Technology Law" 2016; 133: 134–135.
10. Norma ISO/IEC 2382–1: *1993 Information Technology – Vocabulary*, Geneva 1993.
11. Hastie T., Tibshirani R., Friedman J., *The Elements of Statistical Learning. Data Mining*, *Interference and Prediction*, Springer, New York 2008.
12. Frankish K., Ramsey W.M., *The Cambridge Handbook of Artificial Intelligence*, Cambridge University Press, United Kingdom 2014.
13. Bostrom N., *The ethics of artificial intelligence*, w: *Cambridge Handbook of Artificial Intelligence*, Cambridge University Press, United Kingdom 2011.
14. Hastie T., Tibshirani R., Friedman J., *The Elements of Statistical Learning,* Springer, Chicago 2009.
15. Nielsen M.A., *Neural Networks and Deep Learning*, http://neuralnetworksanddeeplearning.com/chap1.html (dostęp: 4.01.2018).
16. Mayer-Schonberger V., Cukier K., *Big Data. A Revolution That Will Transform How We Live, Work and Think*, John Murray Publishers, Chicago 2013
17. Hardesty L., *New Algorithm Lets Autonomous Robots Divvy up Assembly Tasks on the Fly*, Mass. Inst. of Tech., http://www.sciencedaily.com/releases/2015/05/150527142100.html (dostęp: 5.01.2018).
18. Domingos P., Gens R., *Deep Symmetry Networks, Department of Computer Science and Engineering*, Washington 2014, https://homes.cs.washington.edu/~pedrod/papers/nips14.pdf (dostęp: 4.01.2018).
19. Ustawa z dnia 6 listopada 2008 roku o prawach pacjenta i Rzeczniku Praw Pacjenta (Dz. U. z 2017 roku, poz. 1318 ze zm.).
20. Ustawa z dnia 5 grudnia 1996 roku o zawodach lekarza i lekarza dentysty (Dz. U. z 1996 roku, poz. 1318 ze zm.).
21. Karkowska D., *Ustawa o prawach pacjenta i Rzeczniku Praw Pacjenta. Komentarz*. Wyd. Wolters Kluwer, Warszawa 2016.
22. Wyrok SN z dnia 25 marca 1954 roku, sygn. akt II K 174/54.
23. Wyrok SN z dnia 10 lutego 2010 roku, sygn. akt V CSK 287/09, Wyrok SA z Białymstoku z dnia 15 maja 2015 roku, sygn. akt I ACa 1077/14.
24. Wyrok SA we Wrocławiu z dnia 24 stycznia 2014 roku.

25. Ustawa z dnia 10 maja 2010 roku o wyrobach medycznych (uwm) (Dz. U. z 2014 r. poz. 1138, 1162 ze zm.).
26. Rozporządzenie Ministra Zdrowia z dnia 17 lutego 2016 roku w sprawie wymagań zasadniczych oraz procedur oceny zgodności wyrobów medycznych (Dz. U. z 2016 r. poz. 211 ze zm.).
27. Mittelstadt B.D., Allo P., Taddeo M., Wachter S., Floridi L., *The ethics of algorithms. Mapping the Debate*, „Sage Journals" 2016; 2: 1–21.
28. Annany M., *Towards the Ethics of Algorithms Convening, Observation, Probability, and Timeliness*, „Science, Technology and Human Values" 2016; 41 (1): 93–117.
29. Datta A., Sen S., Zick Y., *Algorythmic transparency via quantitative input influence*, Proceedings of 37th IEEE symposium on security and privacy, San Jose, USA, http://www.ieeesecurity.org/TC/SP2016/papers/0824a598.pdf (dostęp: 4.01.2018).
30. Matthias A., *The responsibility gap: Ascribing responsibility for the actions of learning automata*, „Ethics and Information Technology" 2004; 6 (3): 182–183.
31. Art. 6 , art. 20, art. 23 uwm.
32. Bozdag E., *Bias in algorithmic filtering and personalization*, „Ethics and Information Technology" 2015; 15 (3): 209–227.
33. Diakopoulos N., *Accountability in Algorithmic Decision Making*, „Digital Journalism" 2015; 59 (2): 56–62.
34. Hajian S., Domingo-Ferrer J., *A Methodology for Direct and Indirect Discrimination Prevention in Data Mining*, „IEEE Transactions on Knowledge and Data Engineering" 2013; 25 (7): 1445–1459.
35. Hildebrandt M., Koops B.J., *The challenges of ambient law and legal protection in the profiling era*, „Modern Law Review" 2010; 3 (73): 428–460.
36. Newell S., Marabelli M., *Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of 'datification'*, „Journal of Strategic Information Systems" 2015; 24 (1): 3–14.
37. Crawford K., Grey M.L., Miltner K., *Critiquing big data: politics, ethics, epistemology special section introduction*, „International Journal of Communication" 2014; 8: 1663–1672.
38. Geller J.S. L.N., Barash C.I., Billings P.R., Laden V., Natowicz M.R., *Genetic discrimination and screening for hemochromatosis*, „Journal of Public Health Policy" 1994; 15 (3): 345–358.
39. Romei A., Ruggieri S., *A multidisciplinary survey on discrimination analysis*, „The Knowledge Engineering Review" 2014; 29 (5): 582–638.
40. Kamiran F., Karim A., Zhang X., *Decision theory for discrimination-aware classification*, w: Zaki M.J., Siebes A., Yu J.X., Goethals B., Webb G. I., Wu X., *Proceedings of the IEEE International Conference on Data Mining* (ICDM 2012), 2012: 924–929.
41. Kawecki M., *Ogólne rozporządzenie o ochronie danych osobowych. Wybrane zagadnienia*, C.H. Beck, Warszawa 2017.
42. Wyrok Sądu Najwyższego z dnia 3 września 2013 roku, sygn. akt WK 14/13.
43. Wyrok SN z dnia 1 kwietnia 2008 roku, sygn. akt IV KK 381/07.
44. Burrel J., *How the machine 'thinks:' Understanding opacity in machine learning algorithms*, „Big Data & Society" 2016; 3 (1): 1–12.
45. Turilli M., Floridi L., *The ethics of information transparency*, „Ethics and Information Technology" 2009, 11 (2): 105–122.
46. Rozporządzenia Parlamentu Europejskiego i Rady (UE) 2016/679 z 27 kwietnia 2016 roku w sprawie ochrony osób fizycznych w związku z przetwarzaniem ich danych osobowych i w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46.
47. Barta P., Litwinski P., *Rozporządzenie UE w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i swobodnym przepływem takich danych. Komentarz*, Wyd. C.H. Beck, Warszawa 2018.
48. European Commission, *Article 29 Working Party Guidelines on Automated Decision-Making and Profiling for the purposes of Regulation 2016/679, 3 October 2017 r.*, Brussels 2017.
49. Floridi L., Watcher S., Mittelstadt B., *Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation*, „International Data Privacy Law" 2017: 1–47.
50. Wachter S., Mittelstadt B., Russel C., *Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR*, „Harvard Journal of Law and Technology" 2018, https://ssrn.com/abstract=3063289 (dostęp: 27.01.2018).
51. Zielińska E., Sakowski K., Barcikowska-Szydło E., *Ustawa o zawodach lekarza i lekarza dentysty: Komentarz*, Wolters Kluwer Polska, Warszawa 2014.
52. Lisboa P.J.G., *Interpretability in machine learning principles and practice*, „International Workshop on Fuzzy Logic and Applications" 2013: 15–21.
53. Markowetz A., Błaszkiewicz K., Montag C., *Psycho-information: Big Data shaping modern psychometrics*, „Medical Hypotheses" 2014; 82 (4): 405–411.
54. Van Otterlo M., *A machine learning view on profiling*, w: *Privacy, Due Process and the Computational Turn-Philosophers of Law Meet Philosophers of Technology*, Routledge, Abingdon 2013.
55. *Projekt rezolucji w sprawie prywatności w fazie projektowania* – 32. Międzynarodowa Konferencja Rzeczników Ochrony Danych i Prywatności, Jerozolima, 27–29.10.2010 r., http://www. giodo.gov.pl/pl/1520084/3830 (dostęp: 12.12.2017).