Kwartalnik Naukowy Fides et Ratio
Numer 58(2)2024, ISSN 2082-7067

Kompetencje etyczne w uczeniu maszynowym z perspektywy komunikacji w procesie edukacyjnym

Ethical Competencies in Machine Learning from a Communicational Perspective in Educational Process

doi:

Justyna Horbowska[a] ✉

[a] *Mgr Justyna Horbowska, https://orcid.org/0000-0002-0723-0939,*
Szkoła Doktorska KUL, Wydział Filozofii, Katolicki Uniwersytet Lubelski Jana Pawła II
✉ *Autor korespondujący: jhorbowska@kul.lublin.pl*

Abstrakt: Termin „sztuczna inteligencja" (SI) odnosi się do programów komputerowych, wyposażonych w liczne kompetencje, jak dokonywanie obliczeń, grupowanie i kategoryzowanie danych czy komunikowanie się z użytkownikiem w językach etnicznych. Z drugiej strony systemy sztucznej inteligencji nie posiadają pewnych własności, a wśród nich obok braku „zdolności kreatywnych" wyróżniana jest obojętność na aspekt moralny działań przy wyszukiwaniu i kompilacji danych czy katalogowaniu zjawisk. Niniejsze opracowanie ma na celu omówienie wybranych przyczyn takiego stanu rzeczy w kontekście metodyki uczenia maszynowego (ML), z ujęciem problematyki i zastosowań sztucznej inteligencji w perspektywie komunikacji naukowej, jaka zachodzi w procesie edukacyjnym. Wobec tak postawionego celu został sformułowany problem badawczy w postaci pytania: Jak kształtują się kompetencje etyczne w procesie uczenia maszynowego w kontekście komunikacji zachodzącej w procesie edukacyjnym? W celu dokonania próby udzielania odpowiedzi na tak postawione pytanie badawcze posłużono się metodą analizy tekstu, jak również metodą syntezy. W wyniku przeprowadzonych badań ustalono, że prowadzenie uczenia maszynowego z udziałem człowieka, jak również za pomocą systemów sztucznej inteligencji uprzednio uczonych z udziałem człowieka może umożliwiać przekazanie maszynie cybernetycznej treści o charakterze moralnym w rozumieniu normatywnym. Jako, że udział człowieka dopuszcza uczenie nadzorowane maszyn cybernetycznych, ten właśnie rodzaj uczenia, zastosowany jako wyłączna metoda lub w połączeniu z inną metodą niesie możliwość przekazania aplikacjom pożądanych informacji na temat reguł społeczno-kulturowych. W pełni samodzielne uczenie maszyn cybernetycznych nie zapewnia zebrania przez nie informacji dotyczących aspektów etycznych, które są pożądane w komunikacji podczas procesu edukacyjnego, ponieważ otwarte zbiory danych, na których odbywa się uczenie maszyn, mogą zawierać treści szkodliwe, prowadzące do wzmacniania negatywnych zjawisk społecznych.
Słowa kluczowe: sztuczna inteligencja, uczenie maszynowe, kompetencje etyczne, edukacja, komunikacja

Abstract: The term „artificial intelligence" (AI) refers to computer programs equipped with numerous competencies, such as making calculations, grouping and categorizing data, or communicating with the user in ethnic languages. On the other hand, artificial intelligence systems do not have certain properties, and among them, apart from the lack of „creative abilities", is indifference to the moral aspect of actions when searching and compiling data or cataloging phenomena. This study aims to discuss selected reasons for this state of affairs in the context of machine learning (ML) methodology, including the issues and applications of artificial intelligence from the perspective of scientific communication that occurs in the educational process. Given this goal, a research problem was formulated in the form of a question: How are ethical competencies developed in the machine learning process in the context of communication occurring in the educational process? In order to answer this research question, the text analysis method and the synthesis method were used. As a result of the research, it was determined that conducting machine learning with human participation, as well as using artificial intelligence systems previously learned with human participation, may enable the transmission of moral content in the normative sense to a cybernetic machine. Since human participation allows supervised learning of cybernetic machines, this type of learning, used as the sole method or in combination with another method, offers the opportunity to provide applications with the desired information about socio-cultural rules. Fully independent training of cybernetic machines does not ensure they collect information on ethical aspects desirable in communication during the educational process because open data sets on which machine learning takes place may contain harmful content, amplifying negative social phenomena.
Keywords: artificial intelligence, machine learning, ethical competencies, communication

Wprowadzenie

Edukacja w szerokim ujęciu polega na działaniach dotyczących człowieka, a owe działania mają pozostawać w zgodzie z wyznawanym systemem wartości (Okoń, 1998). Również od aplikacji opartych na sztucznej inteligencji (SI) zgodnie z pierwotną definicją McCarthy'ego ujmowanej jako zachowanie maszyny cybernetycznej, które zostałoby uznane za inteligentne w przypadku człowieka (McCarthy, Minsky i in., 1955), oczekuje się umiejętności wartościowania, traktując ją jako kluczową dla rozlicznych zastosowań SI. Przede wszystkim uzyskiwane przez sztuczną inteligencję wyniki mają być adekwatne dla danej dyscypliny, mają mieścić się w jej paradygmacie. O ich poprawności przesądza zgodność z założonymi oczekiwaniami o charakterze formalnym i treściowym. Oprogramowanie do nauki języka ma umożliwić prawidłowe posługiwanie się językiem, a urządzenie służące do przekładów ma dokonać jak najlepszego tłumaczenia tekstu (Massey, Ehrensberger, 2017).

Zatem jeśli ująć to w perspektywie celowej uzyskania wyniku odpowiadającego zapytaniu jako warunku skutecznego komunikowania się (rozumianego tu jako przekazywanie i odbiór komunikatów, które stanowią informację)[1] to podstawowe założenie o nacechowaniu aksjologicznym wydaje się tu spełnione: o dobru jako wartości uzyskanej w efekcie pracy aplikacji przesądzi poprawność wyniku.

Jednak taki rezultat nie odnosi się do dobra moralnego – wynikającego z systemu społecznego, kulturowego, religijnego czy światopoglądowego – ani do powinności moralnej. Jak to określa Hoes: „sztucznej inteligencji brak kompasu moralnego"; dodaje, że jakiegokolwiek innego również (Hoes, 2019). Dzieje się tak dlatego, że systemy sztucznej inteligencji nie są wyposażane w samoświadomość, jak również jej w nich nie stwierdzono (Bishop, 2017). Nie stanowią one również podmiotów komunikacji społecznej nazywanej komunikacją właściwą, kształtowanej przez współintencjonalność jako specyficzną dla ludzkiej kooperacji zdolność do podzielania i współkształtowania celów i intencji (Tomasello, 2022). Status SI może przybliżać argument chińskiego pokoju wprowadzony do debaty filozoficznej przez Searle'a i mający na celu ukazanie w eksperymencie, że maszyna cybernetyczna nawet jeśli przetwarza dane, to nie musi ich rozumieć (Searle, 1980).

Jako, że SI nie posiadają natury, to nie odnosi się do nich *per se* prawo naturalne ani nie obowiązują żadne reguły poza regułami logiki, jeśli nie zostaną im narzucone – za wyjątkiem przypadku uznania praw za równie oczywiste, co obowiązujące w nauce zasady racjonalności teoretycznej (Finnis, 2001).

Wśród wielu podziałów SI[2] zwracają uwagę te, które są oparte na rodzaju zastosowanego uczenia maszynowego. W procesie uczenia szczególną rolę odgrywa źródło pozyskiwanej wiedzy. Pomimo niewątpliwych różnic, proces przyswajania wiedzy przez aplikacje cyfrowe, podobnie jak w przypadku uczniów przebiega na drodze jej pozyskiwania albo poprzez komunikację z podmiotowo traktowanym nauczycielem albo ze środowiska edukacyjnego. Wydaje się, że można wskazać na relacje pomiędzy modelem uczenia algorytmu a możliwością osadzenia pracy SI w kontekście moralnym.

1. Metody uczenia maszynowego

Uczenie maszynowe to zbiór działań zawierających się w zakresie sztucznej inteligencji. Opinie na temat możliwości kreatywnych SI bywają podzielone, jednak przyjmuje się, że uczenie, nawet jeśli samodzielne, w przypadku programów nie zawiera w sobie elementu kreatywności.

Choć zostało wypracowane wiele metod uczenia maszynowego i ich podziałów, można je podzielić na trzy poniższe kategorie:

1 Już inicjator cybernetyki N. Wiener pisał o sposobach przekazywania informacji i porozumiewania się pomiędzy człowiekiem a mechanizmem oraz maszyną a maszyną, (Zob.: N. Wiener, 1961, s. 7).

2 Obok podziałów ze względu na mechanizm uczenia, które zostały tu omówione szerzej istnieją również inne podziały, np. ze względu na sposób dostarczania danych, zastosowania aplikacji czy klasyfikacji systemów uczenia maszynowego. (Przyp. aut.)

1. Uczenie nadzorowane (*ang. supervised learning*),
2. Uczenie nienadzorowane (*ang. unsupervised learning*),
3. Uczenie poprzez wzmacnianie (*ang. RL–reinforcement learning*).

Warto dodać, że każdy z powyższych rodzajów uczenia maszynowego charakteryzuje inny model uczenia się; ponadto każdy z nich jest dedykowany swoistemu rodzajowi problemów, jakie można rozwiązać przy jego pomocy (Flasiński, 2020).

1.1. Uczenie nadzorowane

Do pierwszego rodzaju zaliczane jest nauczanie maszyn cybernetycznych poprzez dostarczanie im danych „oznakowanych"–z dołączonymi gotowymi odpowiedziami na postawiony problem. Odbywa się to w taki sposób, że system otrzymuje przykładowe informacje wraz z ich zaszeregowaniami jako poprawne lub niepoprawne (również np. A lub B), zaś po zgromadzeniu odpowiedniej ilości takich danych i ich analizie jest w stanie samodzielnie szeregować nowe informacje do jednej z grup z poprawnością zależną od ilości wcześniejszych przykładów i jednorodności ich kategorii. W tym przypadku można mówić o „wnioskowaniu" przez analogię, kiedy maszyna odwołuje się do podobnej czy porównywalnej sytuacji orzekając o przyporządkowaniu nowej informacji do jednego ze stanów na podstawie wcześniej nabytej wiedzy.

Uczenie nadzorowane ma na celu wyposażenie maszyny w umiejętność przewidywania wartości lub klasy obiektu. Maszyna osiąga ją na podstawie wcześniejszego przyswojenia dużej liczby przykładów podawanych wraz z etykietami w przypadku wartości lub klasami. Podczas gdy przewidywanie wartości znajduje zastosowanie w rozwiązywaniu problemu regresji, przewidywanie klasy obiektu umożliwia jego odpowiednią klasyfikację (Domingos, 2015).

Jako przykłady algorytmów uczenia nadzorowanego rozwiązujących problem regresji można wymienić:

- regresję liniową,
- regresję wielomianową,
- drzewo regresyjne,
- sieci neuronowe.

Z kolei do przykładów algorytmów uczenia nadzorowanego rozwiązujących problem klasyfikacji są zaliczane:

- drzewa decyzyjne,
- metoda k-najbliższych sąsiadów,
- maszyny wektorów nośnych (SVM – *support vector machine*),
- naiwny klasyfikator Bayesa,
- las losowy,
- regresja logistyczna,
- sieci neuronowe.

Metoda uczenia nadzorowanego pozwala człowiekowi na śledzenie procesów uczenia, ponieważ zakłada kontrolę nad przyswajanymi treściami i nad ich interpretacją w rozumieniu szeregowania na podstawie narzuconych etykiet. Jeśli etykietowanie odbędzie się w oparciu o normy moralne, SI nauczy się na przykładach naśladować rozpoznawanie danych jako dobrych lub złych podobnie, jak to czynią uczniowie na podstawie lektury baśni ludowych pełnych spolaryzowanych moralnie postaci czy też w zaaranżowanej sytuacji edukacyjnej polegającej na spotkaniach z bohaterami, którzy opowiadają o dokonanych szlachetnych czynach.

Ta metoda jest jednak obecnie stosowana dość rzadko, ponieważ cechuje ją czasochłonność i wysoki koszt. Warto tu podkreślić, że choć sieci neuronowe jako aplikacje sztucznej inteligencji mają umożliwiać rozwiązanie obu wzmiankowanych problemów, to ich praca również okazuje się czasochłonna; również koszty sprzętu komputerowego o wielkiej mocy obliczeniowej nie są niskie. Wyższe generacje SI nie wykazują się skutecznością w „nauczeniu" maszyn rozpoznawania dobra i zła porównywalną do człowieka, jeśli same nie zostaną specjalnie przeszkolone w tym obszarze.

1.2. Uczenie nienadzorowane

Ten rodzaj uczenia polega na tym, że do systemu informatycznego trafiają dane bez sugestii co do ich zaszeregowania. Komputer gromadzi je, anali-

zuje, a następnie sam wynajduje elementy wspólne pomiędzy danymi i na ich podstawie łączy dane w grupy. Człowiek pojawia się tutaj dopiero na etapie interpretacji dokonanych podziałów. Ten model może implikować konieczność ograniczenia ilości grup generowanych przez system jako warunek wstępny.

Jeśli chodzi o uczenie nienadzorowane (tak zwane uczenie bez nauczyciela), to cechą, jaka je wyróżnia jest brak etykiet czy klas przypisanych do przyswajanych przez maszynę cybernetyczną przykładów. Znalezienie powiązań między danymi stanowi zadanie do samodzielnego wykonania dla aplikacji. Takie uczenie przygotowuje maszynę do grupowania danych, jak w przypadku algorytmów analizy skupień (ang. clustering). Poza klasteryzacją do grupowania z założeniem jego przyczyny w postaci zależności będą służyły algorytmy wizualizujące nieoznakowane dane w dwóch lub trzech wymiarach (Sala, 2017).

Do najważniejszych z algorytmów wykorzystywanych w uczeniu bez nadzoru – w zależności od problemów, którymi się zajmują – są zaliczane algorytmy służące do analizy skupień, takie jak:

- metoda k-średnich,
- hierarchiczna analiza skupień,
- algorytm grupowania danych (klasteryzacji) oparty na gęstości (DBSCAN–*density-based spatial clustering of applications with noise*) (Starczewski, Goetzen, 2020).

Przykłady algorytmów służących do wyodrębnienia powiązań poprzez wizualizację i redukcję wymiarowości to:

- analiza głównych składowych (PCA – ang. principal component analysis),
- jądrowa analiza głównych składowych.

Jeśli odwrócić problem, metoda uczenia bez nadzoru, otrzymawszy za zadanie wyróżnienie przedmiotów odbiegających od klasterów doprowadzi do wyeliminowania grup i wyłonienia tylko tych rozproszonych elementów, które nie należą do żadnej z nich. Do rozwiązania problemu wykrywania anomalii i nowości może służyć jednoklasowa maszyna wektorów nośnych.

Rezultatem uczenia nienadzorowanego jest zbiór lub zbiory danych zgrupowanych ze względu na jakąś cechę lub cechy wspólne. Przy interpretacji wyników uczenia nienadzorowanego, również w aspekcie etycznym generatywne formy sztucznej inteligencji mogą nie okazać się skuteczne porównywalnie do człowieka dlatego, że człowiek nie ograniczając swobodnego uczenia SI żadnymi warunkami wstępnymi selekcji (w tym informacją na temat moralności) będzie chciał dokonać samodzielnie ostatecznego wyboru z wyłonionych możliwości. W tym modelu nie zachodzi bowiem kwestia „poprawności" lub „niepoprawności" rozwiązania – wartościowanie jest zastąpione przez wyróżnienie równoważnych zbiorów czy elementów. Brak nauczyciela wobec niemającej żadnego wglądu w normy moralne SI w tym wypadku oznacza, że modele w uczeniu nienadzorowanym są całkowicie niewrażliwe moralnie, choć efekty ich pracy nie muszą być moralnie obojętne.

1.3. Uczenie półnadzorowane

Zarówno uczenie nadzorowane, jak nienadzorowane należą do głównych sposobów tradycyjnie wykorzystywanych w uczeniu maszynowym nie tylko odrębnie, ale również jako kombinacja obu metod.

Aby wyeliminować niedogodności związane ze stosowaniem uczenia nadzorowanego, jak jego czasochłonność i kosztowność, a jednocześnie nadać jak najwyższą jakość uczeniu nienadzorowanemu, są konstruowane algorytmy poddawane uczeniu półnadzorowanemu (*ang. semisupervised learning*), z wykorzystaniem danych oznakowanych etykietami lub klasami, choć takie oznakowanie dotyczy zazwyczaj niewielkiej grupy danych. W większości algorytmy półnadzorowane, stanowiące kombinacje algorytmów uczenia nadzorowanego i nienadzorowanego są zaimplementowane do sieci neuronowych.

Innym co do istoty od opisanych wyżej rodzajem uczenia maszynowego jest uczenie przez wzmacnianie.

1.4. Uczenie przez wzmacnianie

Taki rodzaj nauczania zachodzi wówczas, gdy system nie tylko nie otrzymuje przykładowych danych wraz z ich zaszeregowaniem (jako *poprawne/ niepoprawne*, albo *dobre/złe*) jak w uczeniu przez analogię, ale nie uczy się również samodzielnie szeregując dostarczane dane. Ma on za zadanie udzielić odpowiedzi na zapytanie na podstawie przeszukiwania dostępnej bazy danych (często zasobów sieci WWW) bez treningu, niejako *ad hoc* i dopiero reakcja człowieka na zaproponowane przez maszynę odpowiedzi jest źródłem wiedzy dla systemu, który uczy się gdy człowiek wybiera jedną spośród nich zapamiętując skojarzenie pytania z wybraną odpowiedzią jako wzmocnienie czyli sygnał pozytywny. Uczenia maszynowe tego rodzaju odbywa się poprzez interakcję ze środowiskiem na podstawie otrzymywanych zeń na bieżąco informacji – bez wcześniejszego implementowania danych uczących. Pozyskiwanie danych odbywa się automatycznie, a ich oddawanie w reakcji na zapytanie wywołuje reakcję potwierdzenia trafności wyboru (nagroda) lub zaprzeczenia jej. W uczeniu przez wzmacnianie wyróżnia się kluczowe elementy: środowisko, *agenta* i *bufor*.

Przez *środowisko* jest tu rozumiane zadanie (rzeczywiste lub symulacja), z którym wchodzi w reakcję algorytm określany jako *agent* lub *gracz*. Celem uczenia przez wzmacnianie jest maksymalizacja nagrody otrzymywanej przez *agenta* od środowiska, zatem *agent* uczy się osiągania najwyższego możliwego wyniku w danym *środowisku*.

Agent stanowi element, który wchodzi w interakcję ze *środowiskiem* realizując zadanie polegające na nauczeniu się osiągania najwyższego możliwego wyniku, a poprzez to na maksymalizacji uzyskiwanej nagrody. Za zachowanie *agenta* odpowiada funkcja zwracająca akcję, określana jako polityka. Najczęściej polityka jest wprowadzana z zastosowaniem sieci neuronowej.[3]

Z kolei *bufor* to baza danych przechowująca informacje zebrane przez *agenta* podczas uczenia, które następnie służą do jego trenowania.

Jak wynika z powyższego, ten model uczenia może z czasem przyswoić normy kulturowe, religijne czy prawne, do których otrzyma dostęp. Należy jednak pamiętać, że stworzenie odpowiednio szerokiej bazy kontrolowanych danych może przerastać możliwości twórców oprogramowania, a aplikacje otrzymują najczęściej jak najszerszy – niekontrolowany–dostęp do sieci WWW aby mogły czerpiąc z jak największej ilości informacji udzielać skutecznie odpowiedzi na pytania. Można to porównać z szeroko rozumianym środowiskiem edukacyjnym opisywanym tradycyjnie w pedagogice w opozycji do sytuacji edukacyjnej. O ile środowisko edukacyjne otacza ucznia w sposób niekontrolowany, sytuacja „wciąga" go, a jej elementy mają służyć efektowi dydaktycznemu rezerwując miejsce na interpretacje etyczne. Wąsko rozumiane środowisko edukacyjne zakłada autonomię ucznia czerpiącego zeń, bo „już nie nauczyciel naucza z użyciem środków poglądowych, tylko środowisko klasy szkolnej (nie tylko jej) jest bezpośrednim źródłem ważnych impulsów edukacyjnych i konfliktu poznawczego" (Kruk, 2009, s. 494).

Przyrównanie procesu uczenia maszynowego do wypracowanych metod nauczania wychowanków ukazuje, że podobnie, jak w przypadku klasycznego modelu kształcenia, gdzie występują uczeń i nauczyciel, maszyna może mieć mentora, choć nie zawsze będzie to człowiek bo w tej roli może wystąpić sztuczna inteligencja. Uczenie nienadzorowane przywodzi na myśl antypedagogikę–być może uzasadnione byłoby traktowanie tego modelu jak dalekiego echa koncepcji wychowania Rousseau–a uczenie przez wzmacnianie może przypominać tresowanie zwierząt tak, aby nie rozumiejąc przyczyny ani celu, wykonywały określone czynności w reakcji na bodziec z czasem stając się coraz bardziej skuteczne w rozpoznawaniu go.

1.5. Człowiek w procesie uczenia maszynowego

Początkowo w każdym z przybliżonych tu rodzajów uczenia maszynowego udział człowieka był uznawany za niezbędny. Jednak obecne modele sztucznej inte-

3 *Nauka o wzmocnieniu i znaczenie.* (Za:) https://datascience.eu/pl/uczenie-maszynowe/nauka-maszyn-dla-ludzi-czesc-5-nauka-wzmacniania/ (dostęp: 20.01.2024)

ligencji działające w oparciu o sieci neuronowe, jakie rozwinęły się w ciągu ostatnich lat, są programowane w taki sposób, aby uzyskać uprawnienia do zastępowania człowieka w procesie uczenia maszynowego. Oznacza to, że SI jest w stanie uczestniczyć zarówno w uczeniu nadzorowanym, jak nienadzorowanym, a nawet uczeniu poprzez wzmocnienie. Jednak wydaje się, że nie jest w stanie rzeczywiście zastąpić człowieka co najmniej na etapie przyporządkowywania danych do ich etykiet. Również nie można odnieść doń metod nauczania postulowanych w przypadku uczniów, gdy zadanie nauczyciela miałoby sprowadzać się do przybliżania pewnych kontekstów, z których uczeń czerpie jedynie w takim stopniu, jaki uzna za wystarczający (Thomas, Brown, 2011).

Kwestia uczenia maszynowego, szczególnie w kontekście modeli językowych wydaje się nasuwać pytania o relację stosowanych tu kategorii do kryteriów komunikowalności przyjmowanych powszechnie w edukacji i nauce (Kulczycki, 2017). Już w uczeniu nadzorowanym systemów poprawność odpowiedzi oznacza jej zgodność ze zgromadzonym wyborem przykładów, a niepoprawność–brak takiej zgodności. Oznacza to, że poprawność nie musi być tożsama nie tylko z prawdziwością, ale też z jakimkolwiek prawem czy zasadą – stanowi statystyczne prawdopodobieństwo. Uczenie nienadzorowane nie zawiera nawet elementu wartościowania ze względu na stopień poprawności, ponieważ system dzieli elementy zbioru na grupy kierując się wyszukiwanymi przezeń cechami wspólnymi i odczytując je jako równorzędne co do wartości. Uczenie przez wzmacnianie, porównywane z tresurą przywodzi na myśl system kar i nagród, który wydaje się czerpać w jakimś stopniu z behawioryzmu.

Zastosowanie sztucznej inteligencji do usprawniania wyżej opisanych procesów pogłębia zarysowane tendencje do preferowania sądów większości czy wnioskowania na podstawie ilości zjawisk. Upodabnianie sposobu przetwarzania informacji przez SI do mechanizmów działania układu nerwowego przy określeniu takich priorytetów sprawia, że poprawność efektu rozumowań – jakich w takiej sytuacji nie ma możliwości prześledzić – może stanowić dla człowieka nierozwiązywalną zagadkę.

Zatem oczekiwanie potencjalnej przyszłej użyteczności w przypadku uczenia maszynowego pozostaje poza kontekstem moralności, stanowiącym istotny element w komunikacji i edukacji. Zapewne wynika to z faktu, że aplikacje są trenowane aby służyć jako narzędzie. Jednak takie ich ujmowanie również implikuje kolejne wątpliwości, a kwestia wydaje się być bardziej złożona.

Jeśli bowiem maszyna po „wyuczeniu" ma stanowić narzędzie, to kto ma być jego twórcą i użytkownikiem? Narzędzia tradycyjnie były wytwarzane przez człowieka, aby stanowić pomoc dla człowieka, a ich kryterium stanowiła użyteczność. Zatem maszyny uczone poprzez dostarczanie informacji, jakie dobierał człowiek, po wyuczeniu mogły stanowić użyteczne dla niego narzędzie służące również do komunikacji międzyludzkiej. Jednak kiedy narzędzie jest uczone przez SI, jego użyteczność dla człowieka może budzić wątpliwości. Zapewne po wyuczeniu takie narzędzie będzie mogło służyć algorytmowi SI, z kolei SI jest z założenia wykorzystywana przez człowieka. Jednak brak wiedzy co do sposobu, w jaki sztuczna inteligencja uzyskuje rezultaty może budzić wątpliwość co do możliwości ich zastosowania nawet, jeśli będzie je rozpoznawać jako poprawne (Kasperska, 2017).

Z tą sytuacją wiąże się opisywany szeroko w literaturze przedmiotu problem czarnej skrzynki, gdy znane są dane wejściowe i wyjściowe, ale brak informacji co do łączącego je procesu, jaki zachodzi wewnątrz (Chojnowski, 2019). Natomiast skoro człowiek nie jest w stanie w żaden sposób stwierdzić, jakim sposobem rezultat wynikł z przyczyny, to nie może być pewien jego wartości.

Kolejne zagadnienie wynikające z powyższego to algorytmiczne podejmowanie decyzji (ADM). SI dysponuje umiejętnością podejmowania ich i posiłkuje się przy tym pozyskaną wiedzą. Jednak jeśli zgromadzi ją w oparciu o reguły ustalone przez nią samą, to podjęta przez nią decyzja dotycząca człowieka wydaje się nieść niebezpieczeństwo związane z jej konsekwencjami dla człowieka. W tym kontekście jest ujmowane tzw. ryzyko dyskryminacji algorytmicznej (Tolan, 2018).

Choć badania w oparciu o współczesną wiedzę dowodzą braku świadomości i zdolności abstrakcyjnego myślenia zarówno w klasycznej maszynie poddawanej nauczaniu, jak w generatywnej SI, to jednak natura procesów wnioskowania SI opartych na sieciach neuronowych nadal nie jest bardziej czytelna, niż procesów myślowych u człowieka. To tkwiąca u podłoża ludzkiej decyzji struktura światopoglądowa, system wartości urzeczywistniany przez jednostkę decydują o czytelności jej intencji, zamierzeń i czynów. W przypadku SI decyzja może wprawdzie być uzależniona od określonych założeń, jakie sformułował człowiek, jednak brak możliwości prześledzenia procesu warunkuje niewiedzę co do innych założeń poczynionych przez SI choćby na podstawie kryterium największej powtarzalności i wynikającej z niego wyższej skuteczności rozwiązania. Ukazuje to również kwestię odpowiedzialności (współodpowiedzialności) za decyzję takiego rodzaju (Sierocka, 2016); również wtedy, gdy dokonuje ich człowiek w oparciu o rekomendację SI i dotyczą one innych ludzi.

2. Aspekt etyczny stosowania systemów *Big Data*

Specyfika uczenia maszynowego rodzi problemy proporcjonalne do stopnia zaawansowania systemów informatycznych, jakich nie obserwuje się we współczesnej edukacji szkolnej. Jednym z nich jest zagadnienie odpowiedzialności związane z algorytmicznym podejmowaniem decyzji wzmiankowanym wyżej. Sieci neuronowe, aby być efektywne uczą się analizując ogromne zbiory danych (ang. *Big Data*). Choć takie uczenie zapewnia wyższą skuteczność rozwiązań, to jest tracona możliwość dokładnego określenia, jakie czynniki wpłynęły na efekt w postaci dokonania wyboru. Nie można zatem określić, co stanowiło przyczynę w przypadku ewentualnego podjęcia błędnej decyzji.

Kolejnym zagadnieniem jest wzmacnianie uprzedzeń, powielanie treści negatywnych. Kiedy aplikacje uczą się na szerokim spektrum danych, do których mają dostęp, przyswajają również zawarte w nich uprzedzenia czy inne negatywne zjawiska społeczne nie interpretując ich, a jednak powielając i wzmacniając. Przy braku wytycznych odnośnie do uwzględniania choćby systemu wartości obowiązującego w danej kulturze, stanowiącego część kodu kulturowego może to prowadzić do szerzenia dyskryminacji i przemocy. Dlatego szczególnie istotna w procesie uczenia maszynowego systemów trenowanych z myślą o wykorzystaniu w edukacji wydaje się zgodność z założeniami aksjologii pedagogicznej (Maj, 2016).

Gromadzenie danych przez maszyny w trakcie uczenia nie powinno odbywać się z naruszeniem prywatności osób, z których danych aplikacja korzysta. Jednak prywatne dane są przydatne w celu osiągania jak największej skuteczności rozwiązań. W efekcie maszyny potrzebując ogromnej liczby danych mogą wykorzystywać do uczenia wszystkie informacje, do których mają dostęp; mogą je gromadzić, analizować i przetwarzać na różne sposoby bez zgody, a nawet bez wiedzy osób, które je wprowadziły. Niesie to również ryzyko ujawnienia prywatnych informacji o użytkownikach, w tym ich danych osobowych przy udzielaniu odpowiedzi przez maszynę cybernetyczną.

Aby uniknąć wyżej wymienionych problemów, Unia Europejska przyjęła rozporządzenie o nazwie *AI Act*, w którym wyszczególniono zakazane praktyki w stosowaniu SI, jak również wytyczono sposób zatwierdzania oprogramowań obarczonych wysokim ryzykiem. Do praktyk zakazanych należą:

- systemy sztucznej inteligencji wykorzystujące techniki podprogowe, manipulacyjne w celu kształtowania zachowań poszczególnych osób lub społeczności, utrudniające świadome podejmowanie decyzji;
- systemy kategoryzacji biometrycznej na podstawie rasy, poglądów politycznych, przynależności do związków zawodowych, przekonań religijnych lub światopoglądowych, życia seksualnego lub orientacji seksualnej (z wyjątkiem filtrowania sieci w przypadkach łamania prawa);
- systemy SI wykorzystujące informacje na temat wieku, niepełnosprawności, problemów społecznych czy sytuacji ekonomicznej – wykazujące znaczną szkodliwość społeczną;

- systemy sztucznej inteligencji oceniające lub klasyfikujące osoby lub grupy na podstawie zachowań społecznych lub cech osobistych, co prowadzi do szkodliwego lub nieproporcjonalnego traktowania w niepowiązanych kontekstach lub jest nieuzasadnione lub nieproporcjonalne w stosunku do ich zachowania. [4]

Na koniec warto podkreślić, że maszyna cybernetyczna może być podatna na ataki i manipulacje również podczas nauki; gromadzi wtedy błędne dane treningowe. Jeśli nawet samo oprogramowanie aplikacji jest zabezpieczone przed atakami, to może nie spełniać wymogów bezpieczeństwa środowisko, z którego czerpie ona dane. Inną ewentualność stanowi możliwa niezgodność środowiska z oczekiwaniami programisty wynikająca na przykład z zabezpieczenia części potencjalnie potrzebnych danych przez ich właścicieli i w rezultacie liczebnej przewagi w przeszukiwanej przez aplikację sieci WWW treści reklamowych. W takiej sytuacji poprawnie skonstruowana aplikacja zostanie „bezwiednie" wyuczona zwracania błędnych z punktu widzenia jej twórców rozwiązań stając się źródłem zaburzeń w procesie komunikacji.

Podsumowanie

Sztuczna inteligencja może być konstruowana z uwzględnieniem aspektu etycznego, co warunkuje jej wykorzystanie z poszanowaniem człowieka jako osoby w wymiarze społecznym i kulturowym. Ważne jest, aby prowadzić uczenie maszynowe i korzystać ze sztucznej inteligencji w sposób odpowiedzialny, zgodny z normami moralnymi.

SI stanowi narzędzie przydatne w edukacji i może nieść liczne korzyści w aspekcie komunikacyjnym tak dla uczniów, jak nauczycieli czy innych uczestników procesów edukacyjnych, ale jednocześnie rodzi wiele dylematów etycznych i prawnych. Przestrzeganie odpowiednich regulacji i kształtowanie świadomości społecznej są niezbędne, aby zapewnić jej wykorzystywanie w zgodzie z systemem wartości obowiązującym w kulturze. Konieczne jest wyczerpujące informowanie użytkowników o tym, w jaki sposób ich dane są gromadzone, przetwarzane i wykorzystywane przez aplikacje. Należy przestrzegać zakazu wykorzystywania SI do celów szkodliwych dla ludzi i społeczeństwa zawartego w *AI Act* oraz warunków dopuszczania systemów sztucznej inteligencji wysokiego ryzyka.

Bibliografia

Artificial Intelligence Act, European Parliament legislative resolution of 13 March 2024 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonized rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD). (Za:) https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf (dostęp: 15.03.2024).

Bishop, J.M. (2018). Is Anyone Home? A Way to Find Out If AI Has Become Self-Aware, *Frontiers in Robotics and AI, 5*. https://doi.org/10.3389/frobt.2018.00017

Chojnowski, M. (2019). *Zrozumieć decyzje podejmowane przez maszyny*. (Za:) https://www.sztucznainteligencja.org.pl/badacze-z-google-brain-opracowali-system-pozwalajacy-wydobyc-z-modeli-si-informacje-o-stosowanych-kryteriach-oceny/ (dostęp: 15.02.2024).

Domingos, P. (2015). *The master algorithm: How the quest for the ultimate learning machine will remake our world.* New York: Basic Books.

Finnis, J. (2001). *Prawa naturalne i uprawnienia naturalne*. Warszawa: Dom Wydawniczy ABC.

Flasiński, M. (2020). *Wstęp do sztucznej inteligencji.* Warszawa: Wydawnictwo Naukowe PWN.

Hoes, F. (2019). *The Importance of Ethics in Artificial Intelligence*. (Za:) https://towardsdatascience.com/the-importance-of-ethics-in-artificial-intelligence-16af073dedf8 (dostęp: 2.02.2024).

Kamiński, E., *Uczenie maszynowe: z nadzorem i bez nadzoru*. (Za:) https://analityk.edu.pl/uczenie-maszynowe-z-nadzorem-vs-bez-nadzoru/ (dostęp:10.03.2024).

Kasperska, A. (2017). Problemy zastosowania sztucznych sieci neuronalnych w praktyce prawniczej. *Przegląd Prawa Publicznego*, *11*.

Kulczycki, E. (2017). *Komunikacja naukowa w humanistyce*. Poznań: Wyd. IF UAM.

Maj. A. (2016). Aksjologia pedagogiczna. (W:) K. Chałas, A. Maj (red.), *Encyklopedia aksjologii pedagogicznej,* Radom: Polskie Wydawnictwo Encyklopedyczne.

Massey, G., Ehrensberger-Dow, M. (2017). Machine learning: Implications for translator education. *Lebende Sprachen, 62*(2).

4 Zob. *Artificial Intelligence Act, European Parliament legislative resolution of 13 March 2024 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonized rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)*. (Za:) https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf (dostęp: 15.03.2024).

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C.E. (2006). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Magazine, 27*(4), 12. https://doi.org/10.1609/aimag.v27i4.1904

Nauka o wzmocnieniu i znaczenie. (Za:) https://datascience.eu/pl/uczenie-maszynowe/nauka-maszyn-dla-ludzi-czesc-5-nauka-wzmacniania/ (dostęp: 20.01.2024)

Okoń, W. (1998). *Nowy słownik pedagogiczny*. Warszawa: Wydawnictwo Akademickie „Żak".

Sala, K. (2017). Przegląd technik grupowania danych i obszary zastosowań, *Społeczeństwo i Edukacja. Międzynarodowe Studia Humanistyczne*, *2*(25).

Sierocka, B. (2016). Etyka współodpowiedzialności czyli moralność wywiedziona z międzyludzkiej komunikacji. *Rocznik Bezpieczeństwa Międzynarodowego*, *10*(1), 186–196.

Starczewski, A., Goetzen, P., Er, M.J. (2020). A New Method for Automatic Determining of the DBSCAN. *Parameters, Journal of Artificial Intelligence and Soft Computing Research, 10*(3).

Thomas, D., Brown, S.J. (2011). *A New Culture of Learning: Cultivating the Imagination for a World of Constant Change*. Lexington, KY: Creative Space.

Tomasello, M. (2002). *Kulturowe źródła ludzkiego poznawania*. Warszawa: PWN.

Wiener, N. (1961). *Cybernetyka i społeczeństwo*. Warszawa: Wyd. Książka i Wiedza.