

Redakcja naukowa
Artur Wysocki

SZTUCZNA INTELIGENCJA

SZANSE I ZAGROŻENIA

ASPEKTY SPOŁECZNE I ETYCZNE



© Copyright by Uniwersytet Kardynała Stefana Wyszyńskiego
Wydawnictwo Naukowe UKSW, Warszawa 2024

Recenzenci: prof. ucz. dr hab. Mariusz Wojewoda, Uniwersytet Śląski
prof. ucz. dr hab. Joanna Truszkowska, Akademia Łomżyńska

Redakcja i korekta: Anna Postawska

Projekt okładki: Krzysztof Kopania

Projekt typograficzny, skład i łamanie: Renata Witkowska

ISBN: 978-83-8281-452-1 (wersja drukowana)
978-83-8281-451-4 (wersja elektroniczna)

Wydawnictwo Naukowe Uniwersytetu Kardynała Stefana Wyszyńskiego w Warszawie
ul. Dewajtis 5, 01-815 Warszawa, tel. 22 561 89 23, e-mail: wydawnictwo@uksw.edu.pl
www.wydawnictwo.uksw.edu.pl

Druk i oprawa:
volumina.pl Sp. z o.o.
ul. Ks. Witolda 7-9, 71-063 Szczecin

Spis treści

Wprowadzenie	7
I. ZAGADNIENIA ETYCZNE ZWIĄZANE Z ROZWOJEM SZTUCZNEJ INTELIGENCJI	
Ks. ARTUR WYSOCKI Sztuczna inteligencja – narzędzie czy rywal człowieka?	19
LESZEK KORPOROWICZ Sztuczna inteligencja wobec wartości autotelicznych. Pan Twardowski współcześnie	39
AGNIESZKA ZDUNIAK Dylematy etyczne w zastosowaniach sztucznej inteligencji	55
DARIUSZ PROKOPOWICZ Zagrożenia wynikające z nieetycznego i nieuregulowanego prawnie zastosowania technologii sztucznej inteligencji	67
DARIUSZ PROKOPOWICZ Rozwój i zastosowania technologii sztucznej inteligencji z uwzględnieniem aspektów etycznych i społecznych	99
II. WYZWANIA ZWIĄZANE Z ZASTOSOWANIEM SZTUCZNEJ INTELIGENCJI W RÓŻNYCH SFERACH ŻYCIA SPOŁECZNEGO	
SYLWIA JASKUŁA Sztuczny nauczyciel – prawdziwe konsekwencje wychowawcze	131
MAŁGORZATA GRUCHOŁA, ROBERT ZIELIŃSKI Sztuczna inteligencja jako substytut „religii” współczesnego człowieka. Wyniki badań socjologicznych	147

SPIS TREŚCI

AGNIESZKA SZYMAŃSKA	
Sztuczna inteligencja i systemy ekspertowe w diagnostyce i pomocy psychologicznej i psychoterapeutycznej	175
ANNA LESZCZUK-FIEDZIUKIEWICZ	
Rola sztucznej inteligencji w walce z mową nienawiści w mediach społecznościowych na przykładzie Facebooka i Twittera	195
ANNA TURNER	
ChatGPT, BARD, ERNIE – czy algorytmy sztucznej inteligencji to autorytety cyfrowego jutra?	227
HUBERT SIEPERT	
Relacja między człowiekiem a maszyną w kontekście rozwoju sztucznej inteligencji i teologii komunikacji	241

III. SZTUCZNA INTELIGENCJA W MODELARSTWIE SPOŁECZNYM
I TEORII CYBERNETYCZNEJ

MARCIN CHOCZYŃSKI	
Cybernetyczne modelarstwo społeczeństwa przyszłości w perspektywie narzędzia sztucznej inteligencji. ChatGPT a fantastyka i prognostyka socjologiczna	261
STANISŁAW BUDA	
Sztuczna osoba?	287
STANISŁAW BUDA	
Paninteligencja	297

Wprowadzenie

Rozwój techniki i technologii stanowi istotny czynnik rozwoju społecznego i kształtowania dobrostanu człowieka. Człowiek nieustannie doskonali narzędzia służące poprawie życia i wykorzystaniu dóbr otaczającego go świata. Niektóre z wynalazków miały tak duże znaczenie, że zmieniały radykalnie życie i funkcjonowanie całej społeczności ludzkiej: począwszy od opanowania umiejętności rozpalania ognia, posługiwania się nożem, wynalazku koła, stworzenia systemu znaków pisma, aż do bardziej współczesnych, jak maszyna parowa, elektryczność, samolot czy komputer.

Rozwój technologii, którą określamy jako sztuczna inteligencja¹, wpisuje się w cały ciąg dokonań technicznych człowieka zmieniających jego miejsce w świecie, pozwalających lepiej panować nad otaczającym światem, rozwijać go i kształtować. Tak jak każdy istotny moment rozwoju, technologia ta niesie z sobą wiele pytań, wyzwań, szans i zagrożeń. Im potężniejsze narzędzie człowiek zyskuje, tym bardziej szanse, jak i zagrożenia wzrastają. W tym wypadku dążenia sięgają nawet dalej niż tylko stworzenie nowego narzędzia, które ułatwiłoby człowiekowi pracę i pozwoliło mu lepiej żyć. Wraz z tą technologią pojawiają się nowe dążenia do wykreowania tworu podobnego człowiekowi czy może – w pewnym sensie – doskonalszego (roboty humanoidalne, inteligentne maszyny).

¹ Choć zastrzeżenia dotyczące tego określenia, tzn. czy można w tym wypadku mówić o „inteligencji”, należy uznać za uzasadnione, to w niniejszej książce będzie używana ta upowszechniona już nazwa. Jako skrót natomiast stosowany będzie angielski akronim AI (*artificial intelligence*).

Już na tym etapie rozwoju sztucznej inteligencji panuje powszechne przekonanie, że stanowi ona kolejny moment przełomowy w rozwoju ludzkości, niezależnie od tego, czy są to głosy wyrażające zachwyt i wielkie nadzieje, czy wypowiedzi pełne obaw i katastroficznych wizji. Technika jest narzędziem, które można wykorzystać w różny sposób, zarówno do poprawy bytu człowieka, jak i do niszczenia, a zatem ma charakter ambiwalentny. W historii ludzkości wiele doniosłych wynalazków było wykorzystywanych niestety najpierw do walk, zabijania i niszczenia.

Jeśli spojrzymy na współczesny świat, w którym tak niewiele dzieli nas od globalnego konfliktu, i wyobrazimy sobie, że dajemy lepsze narzędzia, większe możliwości agresorom, to nietrudno wyobrazić sobie skutki. W lipcu 2023 r. do kin wszedł film o Robercie Oppenheimerze (*Oppenheimer*), twórcy broni nuklearnej; zostały w nim przypomniane słowa, które bohater zaczerpnął z hinduskiego eposu: „Teraz stałem się Śmiercią, niszczycielem światów”. Energia jądrowa ma kolosalne znaczenie we współczesnym rozwoju, ale wiadomo również, z jakimi tragediami i jakim zagrożeniem się łączy.

Technologia AI rozwija się bardzo szybko w ostatnich latach, a szczególnie wzrost zainteresowania i świadomości społecznej na ten temat wiąże się z udostępnieniem wszystkim internautom od końca 2022 r. systemu ChatGPT przez firmę OpenAI. Od tego czasu uruchomione zostały także przez konkurencyjne firmy, jak Microsoft, Google czy Platforma X (Twitter), inne systemy, które przez interakcję z szeroką publiką i dostęp do olbrzymich baz wiedzy doskonalą swoje możliwości. Samo pojęcie i pierwsze próby zdefiniowania AI oraz koncepcje jej rozwoju powstały w latach 50. XX w.; przyjmuje się, że jako pierwszy użył tego pojęcia John McCarthy w 1956 r.

Pierwsze „laboratoryjne” próby uruchomienia technologii AI przyniosły niestety nie najlepsze wyniki i sprowokowały głosy wzywające do zatrzymania tego procesu. Głosy takie wyrażali nawet ci, którzy obecnie są na czele rozwoju tej technologii. Wydaje się jednak, że możliwość zatrzymania, choćby na jakiś czas, tego procesu jest już nierealna; ze względu na bardzo duży potencjał tej technologii stoją za nim wielkie interesy polityczne i gospodarcze. Trzeba ponadto uwzględnić rozwój omawianej technologii w krajach niedemokratycznych, w których brakuje przejrzystych społecznych reguł działania. Inne państwa nie mogą pozwolić sobie na pozostanie w tyle w tak ważnym momencie światowego rozwoju.

Potrzeba jest zatem szukania właściwych dróg kształtowania i kanalizowania tego rozwoju oraz odpowiedniego edukowania społeczeństwa. Wymagać to będzie nieustannego doprecyzowywania przepisów prawnych, kształtowania kodeksów etycznych, edukacji użytkowników technologii AI, ale także tworzenia systemów bezpieczeństwa, zabezpieczenia czy systemów awaryjnych na wypadek niekontrolowanego rozwoju technologicznego. Aspekty te muszą być doskonalone wraz z nowymi osiągnięciami i możliwościami technologii, a postęp w tym zakresie w związku z samodoskonalącymi się systemami generatywnymi będzie bardzo szybki. Pewne rozwiązania prawne, już zresztą doskonalone, zostały przyjęte w Unii Europejskiej (EU AI Act) czy Stanach Zjednoczonych (w lipcu 2023 r. też wspólna deklaracja największych koncernów dotycząca przestrzegania określonych zasad rozwoju i wykorzystania AI). W Polsce natomiast w drugiej połowie 2023 r. NASK opublikowała raport o cyberbezpieczeństwie w związku z rozwojem AI.

Dla ukształtowania bezpiecznego systemu rozwoju i wykorzystania AI ważne znaczenie ma refleksja naukowa i wymiana doświadczeń między specjalistami zajmującymi się rozwojem nowych technologii. Z jednej strony są to specjaliści badający aspekty techniczne tego rozwoju, z drugiej – specjaliści zajmujący się oddziaływaniem społecznym i jego aspektami etycznymi. Te dwa nurty refleksji i doświadczeń winny się wspierać i dopełniać: społeczne, etyczne i edukacyjne powinny uwzględniać dokonany postęp, a techniczne – brać pod uwagę podnoszone przez etyków, socjologów i pedagogów potencjalne problemy, tak aby w miarę możliwości unikać u początków pułapek niewłaściwego rozwoju.

Niniejsza publikacja stanowi zbiór refleksji autorów różnych dyscyplin nauk społecznych i filozoficzno-etycznych, którzy wskazują na szanse i nadzieje dobrego wykorzystania technologii sztucznej inteligencji, wzmocnienia działań na rzecz dobra wspólnego, a zarazem ostrzegają przed pułapkami niewłaściwego wykorzystania – zagrożeniami, jakie ona niesie. Przy obecnym szybkim rozwoju tej technologii jest jeszcze wiele niejasności i pytań, które również są prezentowane przez poszczególnych autorów. Wszystkie teksty zostały pogrupowane w trzy części. Pierwszy blok prezentuje teksty podejmujące aspekty etyczne i dylematy moralne związane z rozwojem i wykorzystaniem AI, od zagadnień ogólnych do bardziej szczegółowych.

Druga część zawiera teksty opisujące wykorzystanie AI w różnych sferach życia społecznego, w edukacji, sferze religii, psychologii, mediach społecznościowych, nauce i komunikacji. Ostatnia natomiast zawiera interesujące rozważania związane z wykorzystaniem AI w modelowaniu społecznym, a także rozważania z filozofii nauk technicznych z elementami teorii cybernetyki.

W celu lepszego poruszania się po zamieszczonych w publikacji tekstach prezentujemy podsumowania rozważań autorów poszczególnych rozdziałów.

Ks. Artur Wysocki w pierwszym rozdziale, *Sztuczna inteligencja: narzędzie czy rywal człowieka*, rozważa obecne i nadchodzące zmiany społeczne, które mogą się dokonywać pod wpływem rozwoju i wykorzystania technologii sztucznej inteligencji. Po analizie możliwości różnego ukierunkowania rozwoju sztucznej inteligencji Autor podejmuje kwestię zachowania odpowiednich zasad i reguł tegoż rozwoju, tak aby nie były one niszczące dla człowieka i ładu społecznego. Wskazuje na pułapkę pochopnej i nieprzygotowanej akceleracji, która może zachwiać stabilnym funkcjonowaniem społeczeństw. Następnie przedstawiony został zintegrowany system działań, włączający instytucje publiczne, ośrodki badawcze, podmioty gospodarcze i każdego świadomego użytkownika przestrzeni wirtualnej w celu wypracowania podejścia kompleksowego na rzecz właściwego ukierunkowania rozwoju i wykorzystania sztucznej inteligencji.

Leszek Korporowicz, w tekście pt. *Sztuczna inteligencja wobec wartości autotelicznych. Pan Twardowski współcześnie*, wykorzystał znany w literaturze polskiej wątek pana Twardowskiego, który ufając swojej przebiegłości, zaplanował pakt z diabłem, dzięki czemu miał pozyskać wiele dóbr i umiejętności, panując jednocześnie nad warunkami realizacji paktu. W rzeczywistości wpadł jednak w pułapkę i musiał ratować się samotną ucieczką na Księżyc. Wątek ten służy zbudowaniu tezy o analogii wspomnianego paktu z naiwną wiarą udoskonalenia człowieka i jego rzeczywistości przez technologiczny i aplikacyjny postęp sztucznej inteligencji. Tekst jest próbą ukazania stałego dla człowieka pragnienia transgresji, która przynosi tyleż konsekwencji pozytywnych, co negatywnych i kończy się niestety utratą jego podmiotowości, zamienionej na dobrowolną alienację i uzależnienie od zewnętrznych czynników sterujących jego życiem i pojmowaniem samego siebie. Jedną z przyczyn tej porażki jest redukcja ludzkiej osobowości, mitologizująca jedynie fragment

swoich przymiotów osobowości – inteligencję – z pominięciem innych. Redukcja ta spowodowała brak rozróżnienia między inteligencją a mądrością, autonomią a odpowiedzialnością, postępem a rozwojem. W artykule zostają wskazane konsekwencje teoretyczne i praktyczne tego typu redukcjonizmu oraz powiązane z nimi zagrożenia, jakie tkwią w poszukiwaniu post-człowieka oraz post-humanizmu jako ideologii ludzkiego uprzedmiotowienia.

Następny rozdział, przygotowany przez **Agnieszkę Zduniak**, pt. *Dylematy etyczne w zastosowaniach sztucznej inteligencji*, poświęcony jest szansom i zagrożeniom wynikającym z rozwoju i zastosowania sztucznej inteligencji w perspektywie etycznej. Rozwój nowych technologii wpływa w coraz większym stopniu na funkcjonowanie człowieka i społeczeństwa i może mieć też ogromny wpływ na jakość życia przyszłych pokoleń. W tej sytuacji ważna jest refleksja nad etycznym wymiarem sztucznej inteligencji. Główną osią rozważań Autorki jest pojęcie odpowiedzialności w refleksji filozoficznej oraz w jego współczesnym odniesieniu do przemian, jakie zachodzą w społeczeństwie na skutek rozwoju cyfrowych technologii.

W opracowaniu *Zagrożenia wynikające z nieetycznego i nieuregulowanego prawnie zastosowania technologii sztucznej inteligencji* **Dariusz Prokopowicz** zgłębił tytułowy temat, a nadto podkreślił znaczenie etyki odnoszącej się do rozwoju i zastosowań technologii sztucznej inteligencji (AI) w różnych dziedzinach działalności podmiotów gospodarczych i instytucji. Z rozwojem technologii sztucznej inteligencji wiążą się aspekty pozytywne i negatywne, tj. zarówno szanse rozwoju ludzkiej cywilizacji, jak i jego zagrożenia. Z jednej strony rozwój generatywnej sztucznej inteligencji jest swego rodzaju kolejnym etapem dokonującego się już od epoki pierwszej rewolucji przemysłowej postępu technologicznego, polegającego na wzroście skali uprzedmiotowienia pracy i automatyzacji procesów wytwórczych. Z drugiej strony z rozwojem tym wiąże się wiele ryzyk i zagrożeń, których skalę można znacznie ograniczyć, jeżeli na każdym etapie procesu tworzenia, rozwoju i wdrażania technologii AI będą uwzględnione także kwestie etyczne i aspekty prospołeczne, a cały wspomniany proces zostanie normatywnie uregulowany, co powinno w dużym stopniu zmniejszyć skalę prawdopodobieństwa wymknięcia się tego rozwoju spod kontroli człowieka. W rozdziale wskazano na pierwsze potknięcia oraz nieetyczne praktyki biznesowe stosowane w ramach wykorzystania sztucznej inteligencji na portalach internetowych mediów społecznościowych,

a także wynikające z tego zagrożenia. Podkreślona została potrzeba etycznego rozwoju i zastosowań technologii AI.

W kolejnym rozdziale, pt. *Rozwój i zastosowania technologii sztucznej inteligencji z uwzględnieniem aspektów etycznych i społecznych*, **Dariusz Prokopowicz** opisał wybrane istotne kwestie dotyczące rozwoju i zastosowań technologii sztucznej inteligencji z uwzględnieniem aspektów etycznych i społecznych. Z dynamicznie dokonującym się obecnie rozwojem wykorzystania biznesowego technologii generatywnej sztucznej inteligencji wiążą się liczne aspekty natury społecznej i etycznej. Niezbędne jest doskonalenie regulacji prawnych normujących kwestię owego rozwoju i zastosowań, aby nie wymknęły się one spod kontroli człowieka i nie generowały nowych zagrożeń i ryzyk w obszarach ludzkiej działalności. Regulacje prawne dotyczące zastosowań technologii generatywnej sztucznej inteligencji powinny uwzględniać aspekty społeczne i etyczne. W kontekście powyższych rozważań w niniejszym rozdziale Autor zaprezentował sformułowaną przez siebie teorię równowagi w zakresie kwestii przyznawania decyzyjności inteligentnym systemom informatycznym. W kontekście aspektów etycznych omówiono kwestię praw autorskich w odniesieniu do dzieł tworzonych z zastosowaniem generatywnej sztucznej inteligencji. Przedstawiony został również głos papieża Franciszka w temacie przyszłego rozwoju zastosowań sztucznej inteligencji oraz wskazano na konieczność ciągłego dostosowywania uregulowań prawnych do dynamicznie rozwijającej się technologii sztucznej inteligencji.

Sylwia Jaskuła w opracowaniu pt. *Sztuczny nauczyciel – prawdziwe konsekwencje wychowawcze* pokazuje, że AI coraz częściej jest postrzegana jako centralny element cyfrowej transformacji społecznej i cyfryzacji wielu sektorów. Ze wszystkich dziedzin życia, na które będzie miała fundamentalny wpływ, najważniejsza może być edukacja. Sztuczna inteligencja ma potencjał, aby sprostać niektórym z największych wyzwań w dzisiejszej dydaktyce, wprowadzić innowacje w praktykach nauczania i uczenia się, a ostatecznie przyspieszyć postępy w tych zakresach. Rozwijająca się współcześnie AI może znacząco zmienić obecne modele edukacyjne, które mimo stałych, ale nieznaczących ewolucji wciąż zachowują charakter statyczny, odwołując się do dawnych już, mało aktualnych wzorców. W szczególny sposób AI może wspomagać zmiany systemu masowej edukacji w kierunku jego indywidualizacji, choć nie tylko. Z drugiej strony to właśnie edukacja może odegrać znaczącą rolę

w niwelowaniu zagrożeń, jakie już dzisiaj towarzyszą rozwojowi AI. Choć można wskazać wiele zalet wynikających z wdrażania tej technologii do procesów kształcenia, to wykorzystywanie jej w procesach wychowania wzbudza już na obecnym etapie istotne obawy moralne czy społeczne. W dyskusjach dotyczących zastosowania AI w edukacji niepokój dotyczący depersonalizacji i względy etyczne powinny występować na pierwszym planie, tak by została podkreślona potrzeba starannego rozważenia zakresu jej wdrożenia.

W opracowanym wspólnie przez **Małgorzatę Grucholę** i Roberta Zielińskiego tekście pt. *Sztuczna inteligencja jako substytut „religii” współczesnego człowieka. Wyniki badań socjologicznych* została przedstawiona analiza, której celem było pokazanie, że sztuczna inteligencja jest namiastką nowej religii współczesnego człowieka, rozumianej w kategoriach socjologicznych. W rozdziale przedstawiono wyniki ogólnopolskiego badania przeprowadzonego w 2023 r. na próbie 1052 osób w wieku 15–65 lat. W badaniu zastosowano metodę ilościową, technikę kwestionariusza ankiety (panel CAWI), a w doborze próby badawczej – metodę losowo-kwotową. Na początku procesu badawczego przyjęto hipotezę, że sztuczna inteligencja może stać się substytutem religii współczesnego człowieka. Hipoteza została zweryfikowana pozytywnie. Jej potwierdzenie stanowią dane, które dotyczą przedmiotowego i funkcjonalnego rozumienia religii. Choć sztuczna inteligencja staje się substytutem przedmiotu kultu i pełni pewne funkcje religijne, to jednak religią nigdy się nie stanie. Nie uwzględnia bowiem holistycznej koncepcji człowieka, podkreślającej harmonię ducha i ciała. Problem duszy, duchowości jest w przypadku sztucznej inteligencji pomijany.

Agnieszka Szymańska w rozdziale pt. *Sztuczna inteligencja i systemy ekspertowe w diagnostyce i pomocy psychologicznej i psychoterapeutycznej* omówiła problem wykorzystania algorytmów sztucznej inteligencji do udoskonalenia pomiaru psychologicznego, a w szczególności operacjonalizowania definicji probabilistycznych za pomocą logiki rozmytej i oprogramowania ich w systemach ekspertowych. Odniosła się też do wykorzystania cybernetycznego modelu systemu autonomicznego do opisu normy psychologicznej. Autorka przedstawiła również możliwe korzyści, jakie dla sztucznej inteligencji mogą stanowić konsekwencje rozwoju takich dyscyplin, jak: socjologia, pedagogika, filozofia, psychologia. Wydaje się, że mogą one wnieść wkład w ów rozwój przez opisanie w modelach cybernetycznych funkcjonowania normy

psychologicznej oraz poprawnego funkcjonowania społecznego. Taki opis mógłby zostać wykorzystany do rozwoju programów sztucznej inteligencji. Mógłby ją uczynić bardziej podobną do człowieka nie tylko w zakresie jego myślenia (jak miały to za zadanie uczynić algorytmy sztucznej inteligencji), lecz także w zakresie przeżywania przez niego uczuć (nawet „sztucznych uczuć”) oraz wartości i norm społecznych.

W opracowaniu *Rola sztucznej inteligencji w walce z mową nienawiści w mediach społecznościowych na przykładzie Facebooka i Twittera* **Anna Leszczuk-Fiedziukiewicz** przedstawiła analizę polityki właścicieli mediów społecznościowych wobec kwestii stosowania sztucznej inteligencji na przykładzie dwóch gigantów technologicznych, Facebooka i Twittera. Omówiła zarówno ich postulaty, inicjatywy, jak i liczne niekonsekwencje w traktowaniu tzw. szkodliwych treści (ang. *harmful content*), postrzeganych jako mowa nienawiści. Badaczka wykorzystwała metodę analizy danych zastanych, korzystając z publikacji obnażających kulisy pracy Facebooka, i analizę dyskursu medialnego skoncentrowanego na eksperckiej ocenie działań monopolistów. Konkluzje artykułu koncentrują się na skomplikowanej roli, jaką sztuczna inteligencja odgrywa w zarządzaniu treściami i komunikacją, ponieważ Facebook i Twitter zapewniają opinię publiczną o wykorzystywaniu algorytmów do eliminowania materiałów o wydźwięku rasistowskim, obraźliwym, jednocześnie realizując swoje cele biznesowe i marketingowe. Wykorzystywane przez nich algorytmy pełnią funkcję miecza obosiecznego, napędzając ruch i zaangażowanie w sieci oraz czyszcząc internet z oburzających wpisów. Niezwykle istotna jest presja społeczeństwa i współpraca z instytucjami, by uczenie maszynowe wykorzystywane do usuwania mowy nienawiści miało charakter etyczny i odpowiedzialny.

Anna Turner w rozdziale pt. *ChatGPT, BARD, ERNIE – czy algorytmy sztucznej inteligencji to autorytety cyfrowego jutra?* wskazuje, że cyberprzestrzeń niepostrzeżenie stała się kolejnym wymiarem życia społecznego i kulturalnego, w którym użytkownicy sieci pracują, spędzają czas wolny, korzystają z dóbr kultury i poszukują kontaktów społecznych. Wraz z rozwojem sieci korzystanie z niej staje się coraz łatwiejsze i wygodniejsze, ale pojawiają się również pokusy i zagrożenia, które wywołują dyskusje na temat zakresu i jakości jej wpływu na nasze życie i zachowanie. Autorka skupiła się w tekście na trzech aspektach: konsekwencjach nieuchronnych przeobrażeń kultury

eksperskiej środowiska naukowego, oddziaływaniu nowych technologii na funkcjonowanie wirtualnej przestrzeni wpływu, a także na roli chatbotów – algorytmów sztucznej inteligencji, które mogą diametralnie zmienić sposób, w jaki wchodzimy w interakcję z technologią. Zwraca ona również uwagę na ludzi wykluczonych informacyjnie i cyfrowo.

Hubert Siepert w rozdziale pt. *Relacja między człowiekiem a maszyną w kontekście rozwoju sztucznej inteligencji i teologii komunikacji* przedstawia perspektywę teologiczną w ocenie rozwoju sztucznej inteligencji i środków komunikacji. Przypomina, że ostatnie 250 lat historii ludzkości to okres ciągłego rozwoju technologicznego, który pociąga za sobą przemiany na poziomie społecznym i indywidualnym. Aktualnie na horyzoncie pojawia się kolejna przemiana cywilizacyjna, powiązana z rozwijającymi się nowymi technologiami teleinformatycznymi oraz systemami sztucznej inteligencji. Owym zagadnieniem powinien zająć się także Kościół, który szczególnie współcześnie ma być rzeczywistością wskazującą na duchowy wymiar człowieka. Z tego też względu w niniejszym artykule następuje próba przedstawienia nowych wyzwań związanych z relacjami, jakie pojawiają się między człowiekiem a systemami AI w świetle teologicznej refleksji Kościoła. Na początku tekstu Autor ukazuje pewne naturalne predyspozycje człowieka do życia społecznego i relacyjności w świetle nauk ewolucyjnych i neurologicznych. Potem przedstawia rozwój systemów sztucznej inteligencji, ze szczególnym uwzględnieniem naturalnego przetwarzania języka przez te systemy, co pociąga za sobą coraz bardziej intuicyjny i bardziej „ludzki” kontakt z owymi twórcami technologicznymi. Na koniec przedstawia perspektywę teologiczną oceny możliwości i jakości relacji między człowiekiem a systemami AI.

Kolejny tekst, pt. *Cybernetyczne modelarstwo społeczeństwa przyszłości w perspektywie narzędzia sztucznej inteligencji. ChatGPT a fantastyka i prognostyka socjologiczna*, opracowany przez **Marcina Choczyńskiego**, porusza problematykę oczekiwań społecznych, przetestowaną realistycznie w serii pytań zadanych sztucznej inteligencji – w ogólnodostępnym programie komputerowym ChatGPT w wersji 3.5. Jest to tzw. chatbot, który działa na zasadzie czatu, formułując odpowiedzi na pytania zadawane przez użytkowników. Odpowiedzi te pochodzą z przetwarzania informacji pobieranych przez algorytmy oprogramowania z internetu. Pytania na ogół dotyczyły istoty i natury przyszłego społeczeństwa, w tym jego modelowania (strukturyzacji),

ale ich specyfika jest zróżnicowana. Wcześniej zaprezentowane zostały wybrane informacje na temat antycypacji społecznej, zarówno z naukowego punktu widzenia prognozowania socjologicznego, jak i z literackiej wizji fikcji – w tzw. fantastyce socjologicznej, która jest podgatunkiem science fiction. Odpowiedzi uzyskane z ChatGPT zostały porównane zarówno z ustaleniami naukowymi, jak i dziełami literackimi.

Stanisław Buda w pierwszym swoim tekście, pt. *Paninteligencja*, rozważa problem stopniowego przeradzenia się uniwersum całkowicie pozbawionego inteligencji w uniwersum, o którym można rzec, że jest inteligentnym podmiotem, osobą. Podmiot może obejmować albo wyłącznie jego system decyzyjny (umysł, system przetwarzający dane), albo całość, na którą składają się zarówno system decyzyjny, jak i system wykonawczy, przez który należałoby rozumieć mechanizm umożliwiający jego funkcjonowanie. Autora interesuje kwestia, czy środowisko zewnętrzne względem inteligentnego podmiotu może być stopniowo zmieniane i włączane w obręb jego podmiotowości, a dokładniej: w obręb systemu wykonawczego jego systemu decyzyjnego. I dalej: czy ostatecznie zakres jego systemu wykonawczego może pokryć się z zakresem uniwersum. Wymagałoby to od takiego podmiotu wiedzy o sobie samym, dotyczącej z jednej strony własnego wnętrza (w tej terminologii chodziłoby o system wykonawczy), z drugiej zaś – sięgających poza to wnętrze hipotez umożliwiających całościową wizję samego siebie. Wiedza dotycząca własnego wnętrza nie może być pełna, a realny jej przyrost może być oceniany wyłącznie z perspektywy „zewnętrznej”, w świetle opartej na hipotezycznych założeniach (metafizycznej) autointerpretacji. Owa „metafizyka” jest też niezbędną przesłanką motywacji do podjęcia jakiegokolwiek w pełni autonomicznej i autotelicznej aktywności.

W kolejnym zamieszczonym w książce rozdziale, pt. *Sztuczna osoba*, **Stanisław Buda** dzieli się natomiast kilkoma refleksjami dotyczącymi mechanizmów działania człowieka, systemu inteligentnego, inteligentnego agenta oraz podmiotu superinteligentnego. W tym kontekście definiuje takie pojęcia jak: podmiot, autonomiczny system, inteligencja, środowisko, świadomość, samoświadomość, autoteliczność, osoba. Rozważa możliwość autonomicznego działania sztucznego podmiotu. Twierdzi, że podmiot taki byłby osobą tylko wtedy, gdyby był źródłem własnych intencji. Wówczas jednak musiałby filozofować, czyli operować kategorią zarówno sensu, jak i sensowności.

I. Zagadnienia etyczne związane z rozwojem sztucznej inteligencji

Copyright by Wydawnictwo Naukowe UKSW

Sztuczna inteligencja – narzędzie czy rywal człowieka?

Zmiana technologiczna związana z wykorzystaniem sztucznej inteligencji będzie skutkowała istotnymi zmianami społecznymi, które aby zachować ład społeczny, wymagają poważnej refleksji. Podstawą każdego ładu jest system wartości i norm zachowywanych przez członków społeczeństwa, określający reguły i zasady postępowania i działania mające ochronić przed szkodą wyrażoną innemu człowiekowi czy całym społecznościom. Nie tylko jednak stabilny ład wymaga takiego systemu prawno-etycznego, ale również czas przejścia morfogenetycznego prowadzący od jednego układu społecznego do innego. Winien on także dokonywać się według pewnych reguł i zasad, tak aby nie miał charakteru „wyrotowej” rewolucji, która następnie „pożera własne dzieci”². Jeśli socjologowie, jak Margaret Archer, uważali ostatnie dwa wieki za okres społeczeństwa morfogenetycznego – w trakcie ciągłych zmian³, to obecnie, w związku z nowymi technologiami, wiele zmian będzie jeszcze przyspieszać. Czy jednak człowiek lub całe społeczeństwa są zdolne do tak szybkich adaptacji? Już teraz można usłyszeć głosy mówiące, że wiele

¹ Ks. dr hab. Artur Wysocki, prof. ucz., Uniwersytet Kardynała Stefana Wyszyńskiego w Warszawie

² Por. P. Sztompka, *Socjologia zmian społecznych*, tłum. J. Konieczny, Znak, Kraków 2010, s. 273–296.

³ D. V. Porpora, *Morphogenesis and Social Change*, w: M. Archer (ed.), *Social Morphogenesis*, Springer, London–New York 2013, s. 35.

rzeczy dzieje się za szybko, że ludzie nie nadążają, że to ich przerasta, i można zaobserwować narastający trend, w tym również wśród osób młodych, do usuwania się na różne ubocza życia społecznego w poszukiwaniu równowagi i harmonii życiowej. Czy technologia AI pogłębi te problemy? czy pozwoli je złagodzić? a może wprowadzi zmianę, która zmarginalizuje cały gatunek ludzki, gdyż takie zapowiedzi również słychać – jak przestrzegał chociażby Yuval N. Harari⁴?

Czy należy obawiać się maszyn?

Przed rozważaniem naszego stosunku do AI i sformułowaniem pewnych zaleceń związanych z jej rozwojem i wykorzystaniem potrzeba sobie uświadomić, z jaką rzeczywistością mamy do czynienia. Począwszy od pierwszej definicji Johna McCarthy'ego w 1956 r. powstało już wiele określeń tego, co oznacza sztuczna inteligencja, nie ulega jednak wątpliwości, że będą one ewoluowały wraz z tworzeniem nowych możliwości i zastosowań. Najogólniej rzecz ujmując, obecnie wyróżnia się proste systemy zadaniowe, AI pierwszej generacji, która nazywana jest słabą AI, oraz systemy ogólne oparte na uczeniu maszynowym, samodoskonalące się. Jest to AI drugiej generacji, nazywana silną AI⁵ czy obecnie najczęściej AI generatywną (GenAI). Wyróżnia się jednak wiele modeli GenAI, w zależności od zastosowanych modeli głębokiego uczenia się (*deep learning*), wykorzystania rozmaitej architektury sieci neuronowych (*neural network*) itd.⁶

Rozwój generatywnej AI daje już trudne do wyobrażenia możliwości jej zastosowania na potrzeby człowieka i życia społecznego, niemniej jednak jest to dopiero początek dalszych zastosowań omawianej technologii. Kolejna

⁴ Ch. Taylor, 'Sapiens' author says AI is an alien threat that could wipe us out: 'Instead of coming from outer space, it's coming from California', 12.09.2023, <https://fortune.com/2023/09/12/sapiens-author-yuval-noah-harari-ai-alien-threat-wipe-out-humanity-clon-musk-steve-wozniak-risk-cogx-festival/> (5.02.2024).

⁵ J. R. Searle, *Minds, brains, and programs*, „The Behavioral and Brain Sciences” 1980, vol. 3, s. 417–418.

⁶ S. Feuerriegel, J. Hartmann, Ch. Janiesch, P. Zschech, *Generative AI*, „Business & Information Systems Engineering” maj 2023, doi.org/10.1007/s12599-023-00834-7.

istotna zmiana wiąże się z połączeniem tejże technologii z odpowiednim modułem wykonawczym. Oznacza to tworzenie mniej lub bardziej autonomicznych jednostek, które będą mogły realizować działania, a nawet całe ich procesy bez polecenia ze strony człowieka⁷. Stworzenie modułu wykonawczego upodobnionego do ciała ludzkiego jest już w wielu ośrodkach badawczych na zaawansowanym etapie prac. Roboty humanoidalne przy odpowiednim zaawansowaniu mogą wykonywać bardzo wiele działań realizowanych przez ludzi, a do tego być utożsamiane z człowiekiem, lub ich działania mogą być interpretowane jako działania ludzkie. Systematyzację upodobniania się AI do człowieka w przejrzysty sposób przedstawili już S. J. Russell i P. Norvig w swojej książce *Artificial Intelligence*⁸.

Wykorzystanie technologii AI zmierza także równolegle w kierunku, który wydaje się rodzić nie tylko największe obawy o jego skutki dla ludzkości, ale również największe dylematy etyczne. Dotyczy to wykorzystania AI w tworzeniu hybryd zwierzęta-maszyny (*animal-machine hybrids*) i ludzkie-maszyny (*human-machine hybrids*). Nie jest to też już kwestia przyszłości, jesteśmy bowiem świadkami kolejnych etapów rozwoju tego typu eksperymentów. Nie tylko istnieją już różne hybrydy zwierzę-maszyna⁹, ale dokonuje się również rozmnażania tego typu jednostek¹⁰. Jeszcze bardziej dyskusyjny jest rozwój wykorzystania technologii w tworzeniu bytów hybrydowych z organizmem ludzkim. Funkcjonują już przewodniki służące postępowaniom w tego typu procedurach¹¹.

⁷ M. Ciesielski, *Generatywna AI to dopiero początek. Nadchodzą autonomiczni agenci*, 11.01.2024, <https://forsal.pl/lifestyle/technologie/artykuly/9398795,generatywna-ai-to-dopiero-poczatek-nadchodza-autonomiczni-agenci.html> (7.02.2024).

⁸ Zob. J. Russell, P. Norvig, *Artificial Intelligence. A Modern Approach*, Pearson, New Jersey 2010. Por. M. Ferdynus, *Czy mamy prawo „oswoić” sztuczną inteligencję*, „Ethos” 2023, t. 36, nr 143(3), s. 67–91.

⁹ Np. T. Thang Vo-Doan, H. Sato, *Insect-machine Hybrid System: Remote Radio Control of a Freely Flying Beetle (*Mercynorrhina torquata*)*, „Journal of Visualized Experiments” September 2016, doi: 10.3791/54260.

¹⁰ S. Jaba, *World's First Animal-Machine Hybrids Start to Reproduce*, 17.12.2021, <https://chapterzmagazine.com/2021/12/worlds-first-animal-machine-hybrids-start-to-reproduce/> (8.02.2024).

¹¹ G. Demartini, D. E. Difallah, U. Gadiraju, M. Catasta, *An Introduction to Hybrid Human-Machine Information Systems*, „Foundations and Trends” in Web Science” 2016, vol. 7, no. 1, s. 1–87. Zob. również: G. Demartini, D. E. Difallah, U. Gadiraju, M. Catasta, *An Introduction to Hybrid Human-Machine Information Systems*, Now Foundations and Trends, Boston 2017.

Rozwój ten uzasadniany jest najczęściej dobrem społeczności ludzkiej, nowymi możliwościami w medycynie, w kwestii bezpieczeństwa i tak dalej, tak jak to się dzieje w przypadku tworzenia hybryd zwierzęco-ludzkich w wielu instytutach badawczych na świecie. Pierwszym etapem w tym względzie było tzw. *human enhancement*, czyli próba udoskonalania człowieka, służąca naprawieniu jego braków lub wzmocnieniu obecnych możliwości¹². Wykorzystanie AI w tym względzie otwiera bez wątpienia nowe perspektywy i możliwości. Powstają tu jednak dylematy dotyczące tego, jak daleko można ingerować w organizm ludzki, czy jest to rzeczywiście dobre dla człowieka i nie spowoduje chociażby zachwiania jego tożsamości.

Dylematy te stają się jeszcze bardziej pogłębione w przypadku tworzonych hybryd, określanych też jako cyborgi, czyli świadomego łączenia zaawansowanej technologii z organizmem ludzkim, w tym przypadku należy zapytać, czym są lub czym będą i czego można od nich oczekiwać. O rozwoju koncepcji cyborgizacji i współczesnych próbach jej realizacji pisze m.in. Danuta Smołucha w tekście *Cyborg – mariaż człowieka i technologii*¹³.

Wykorzystanie organizmów ludzkich (ludzi?) czy zwierzęcych (zwierząt?) do tworzenia hybryd wynika z faktu ograniczonych możliwości człowieka, mimo dokonanego – wydaje się – tak wielkiego postępu. Człowiek nie jest w stanie stworzyć życia, nawet w jego najmniejszej i najprostszej formie, np. żdźbła trawy. Może on wykorzystywać byty żyjące, manipulować nimi, ich życiem, organizmami, naśladować ich konstrukcję, „rozwiązania techniczne”, ale nie jest dawcą życia ani jego panem. W tym obszarze, w którym człowiek chce dorównać Stwórcy, rodzi się wiele wyzwań i problemów etycznych. Czy wykreowane byty to maszyny? czy ludzie? czy zwierzęta? Jak je traktować? Jakie przyznać im prawa? Wydaje się, że czasami na te pytania nie będzie jednej prostej odpowiedzi. Ponadto rodzi to realne zagrożenie stworzenia bytu czy bytów, które będą niszczycielskie dla człowieka.

¹² Por. A. Alichniewicz, M. Michałowska, *Udoskonalanie gatunku ludzkiego. Problemy definicyjne i argumentacyjne*, „Principia” 2018, nr 64, s. 231–261; D. Cebo, *Human enhancement*, „SSRN Electronic Journal” styczeń 2020, doi: 10.2139/ssrn.4170853.

¹³ D. Smołucha, *Cyborg – mariaż człowieka i technologii*, „Perspektywy Kultury” 2023, nr 42(3), s. 191–199.

Należy pamiętać, że nie tylko sama możliwość techniczna czy technologiczna powinna ludziom pozwalać czynić coś, co jest pokusą zwłaszcza przy nowych narzędziach, ale pierwsze winny być pytania: po co to czynić? czemu to ma służyć? czy rzeczywiście będzie to czymś dobrym dla człowieka i całej ludzkości? Warto w tym miejscu przytoczyć słowa Benedykta XVI na ten temat:

Rozwój technologiczny może zrodzić ideę samowystarczalności techniki, jeżeli człowiek, stawiając sobie tylko pytanie jak, nie zastanawia się nad wieloma dlaczego, które skłaniają go do działania. (...) mentalność technicystyczna staje się tak mocna, że prowadzi do utożsamienia prawdy z rzeczą możliwą do zrobienia. Ale kiedy jedynym kryterium prawdy jest skuteczność i użyteczność, rozwój zostaje automatycznie zanegowany. Prawdziwy bowiem postęp nie polega w pierwszym rzędzie na robieniu czegoś. Kluczem do postępu jest inteligencja zdolna do myślenia o technice i zrozumienia w pełni ludzkiego sensu działania człowieka, w perspektywie sensu osoby wziętej w całości swego bytu¹⁴.

Pułapka akceleracji

Przyspieszenie, jakie niesie z sobą rozwój i wykorzystanie nowych technologii, dla wielu ludzi może być nieznośne i nie tylko rodzić zamieszanie, niepewność i niepokój wewnętrzny, które, jak wskazuje psychologia, oddziałują destrukcyjnie, ale także wzmacniać społeczne niepokoje. Współczesny człowiek, ogarnięty gorączką postępu, wydaje się mieć wszystko z wyjątkiem czasu. Jest paradoksem, że maszyny i urządzenia, które miały wyręczać człowieka w pracy, dać mu więcej czasu wolnego, sprawiają, że jest on przez nie pochłaniany: ich rozwojem, wykorzystaniem i troską o nie. Dotyczy to nade wszystko człowieka świata zachodniego.

¹⁴ Benedykt XVI, *Caritas in veritate. O integralnym rozwoju ludzkim w miłości i prawdzie*, Watykan 2009, nr 70.

Nie jest to oczywiście „wina” maszyn i urzędów, ale człowieka, który pośród tak wielu nowych wynalazków nie potrafi właściwie pokierować swoim rozwojem, gubi się i traci to, od czego zależy jego integralny rozwój, zachowanie tożsamości, prawdziwe dobro. Człowiek zdaje się nieprzygotowany, brak mu odpowiedniej edukacji do wykorzystania nowych osiągnięć, ale też w znacznym stopniu sam odrzucił to, co decydowało o jego indywidualnej stabilności emocjonalnej i ładzie społecznym, ulegając zarazem procesom sterowania politycznego, gospodarczego i kulturowego. Fundamenty związane z podstawowymi grupami społecznymi, przekazywanymi przez nie wartościami, począwszy od rodziny przez Kościoły i wspólnoty religijne, a na różnych wspólnotach i stowarzyszeniach społeczności lokalnych kończąc, zostały w znacznym stopniu zapoznane. Zachłyśnięcie się postępowaniem daje poczucie wielkich możliwości samorealizacji bez oglądania się na to, co tworzyło podstawy stabilnego ładu społecznego.

Dlatego potrzeba powrotu do podstawowych pytań, zwłaszcza w obliczu oczekiwanych głębokich przemian społecznych, jakie przyniesie dalszy rozwój i wykorzystanie sztucznej inteligencji. Należy przypomnieć pytania związane z rozwojem: czy wszystko może się zmieniać i czy w takim samym tempie? Na pewno nie mogą się zagubić fundamentalne zasady życia społecznego, stanowiące o możliwości przetrwania i normalnego rozwoju społeczności ludzkiej, jak szacunek dla godności każdego człowieka i każdego narodu, powszechne przeznaczenie dóbr, troska o dobro wspólne itd. Wiele wartości i norm kulturowych wytworzonych przez poszczególne społeczności, choć może ulegać zmianom, to jednak dla zachowania stabilności systemu wymaga czasu, ponieważ muszą one zostać stopniowo przyswojone i zinternalizowane przynajmniej przez większość społeczności.

Szybko rozwija się technika i jej zastosowanie, zwłaszcza że najczęściej łączą się z tym interesy określonych grup. Nie ulega jednak wątpliwości, że jest ona pomocna w realizacji wielu zadań, które podejmuje człowiek. Z rozprzestrzenieniem się i upowszechnieniem technologii łączy się również dążenie nazywane od pewnego czasu demokratyzacją technologii¹⁵, mające sprawić, że nowe technologie będą coraz łatwiej dostępne, możliwe do

¹⁵ Zob. A. Feenberg, *Democratizing Technology: Interests, Codes, Rights*, „The Journal of Ethics” June 2001, vol. 5(2), s. 177–195.

obsłużenia przez każdego zwykłego człowieka, bez konieczności posiadania wiedzy specjalistycznej. Bardzo dobrze widać to w przypadku rozmaitych urządzeń elektronicznych, począwszy od komputera osobistego. Przyspieszenie spowodowane przez rozwój AI będzie jeszcze większe. I w tym wypadku od wielu lat mówi się o demokratyzacji tej technologii – jako jeden ze strategicznych trendów technologicznych wskazuje ją od kilku lat grupa analityków Gartner Research¹⁶. Rozwój AI w postaci generatywnej (GenAI) ma z kolei istotnie poszerzyć zakres demokratyzacji wiedzy i umiejętności¹⁷. Demokratyzacja AI i udostępnienie jej do powszechnego użytku wiąże się w tym wypadku nie tylko z możliwością korzystania z łatwego dostępu do niej dla przeciętnego człowieka, ale również z możliwością szybkiego rozwoju samej technologii, która w kontaktach z tak dużą liczbą osób doskonali swoje „umiejętności”.

Ten powszechny łatwy dostęp może jednak również być przestrzenią sterowania społecznego, którego uczestnicy mogą nie być świadomi interakcji z systemami AI. W pewnym sensie można to porównać z wykorzystaniem do interesów politycznych czy gospodarczych tzw. kultury masowej, którą m.in. z tej racji Krzysztof Wielecki zalicza do nurtu antykultury, oddziałującej negatywnie na podmiotowość człowieka¹⁸. Sztuczna inteligencja niesie zagrożenie nowymi formami dominacji i władzy, które mogą te procesy jeszcze pogłębić¹⁹. W jakim stopniu człowiek jest przygotowany na jej wykorzystanie tak, aby nie została zredukowana jego podmiotowość, lecz by technologia służyła jego integralnemu rozwojowi? Czy jest w stanie właściwie „udomowić” AI, co rozważa m.in. Mariusz Wojewoda w swoim tekście *Artificial Intelligence as a Social Utopia*? Odwołując się do myśli współczesnych autorów traktujących o rozwoju tej technologii, prezentuje on formułowane

¹⁶ Gartner Research, *Predicts 2019: The Democratization of AI*, 29 listopada 2018, <https://www.gartner.com/en/documents/3894386> (7.02.2024).

¹⁷ Gartner Research, *Generative AI Can Democratize Access to Knowledge and Skills*, 17 października 2023, <https://www.gartner.com/en/articles/generative-ai-can-democratize-access-to-knowledge-and-skills> (7.02.2024).

¹⁸ K. Wielecki, *Kultura versus kultura masowa*, NCK, Warszawa 2024, zwłaszcza s. 535–608.

¹⁹ Franciszek, *Sztuczna inteligencja i mądrość serca: dla komunikacji w pełni ludzkiej*, 24.01.2024, *Orędzie* na 58. Światowy Dzień Środków Społecznego Przekazu 2024, <https://www.vatican.va/content/francesco/pl/messages/communications/documents/20240124-messaggio-comunicazioni-sociali.html> (5.02.2024).

przez nich ostrzeżenia przed brakiem władzy nad AI, co rodzić będzie bardzo realne zagrożenia, według niektórych – tragiczne w skutkach dla całej ludzkości²⁰. Rozwinięte systemy AI mogą stopniowo przejmować kontrolę nad człowiekiem, nawet nad jego wyobraźnią i planami²¹.

Technika będąca wytworem człowieka, wykorzystywana przez niego do konkretnych działań i ukierunkowywana – jak w przypadku AI – traci swój neutralny charakter²². Jej działania są efektem zarówno przekazanego intencjonalnie programu, elementów nieuwzględnionych w jego funkcjonowaniu, jak i pozwolenia na jej autonomiczny rozwój dzięki dostępowi do szerokiej wiedzy, jaką daje przestrzeń wirtualna: systemy AI będą zdobywały wszelką dostępną tam wiedzę, z czego trzeba zdawać sobie sprawę. Maszyny nie będą miały świadomości i odpowiedzialności w sensie ścisłym, chyba że będą to byty hybrydowe. Czy mogą być twórcze? Mogą na pewno, operując szeroko dostępną wiedzą i dokonując pewnych przekształceń, używając zasad probabilistyki itd., tworzyć nowe rzeczy, w tym także coś, co można by określać jako dzieła sztuki. Wystarczy również, że AI „nakarmi się” wiedzą, o której nie wiemy, a jej działania mogą się okazać nieprzewidywalne, i w tym sensie „kreatywne”. Dodatkowe wprowadzenie probabilistyki do jej działań w innych sferach może w odbiorze upodobić jej postępowanie do ludzkiego, w sensie nieprzewidywalności, ale wszyscy zdają sobie sprawę, co może to oznaczać w przypadku „bezdusznych maszyn” – możliwości działania nakierowanego przeciwko człowiekowi gwałtownie rosną.

Rozwój technologii AI jest związany z dalszym poznawaniem człowieka i jego godności, tego, co go odróżnia od innych stworzeń, a tym bardziej od nawet najlepszych maszyn. Leszek Korporowicz pisał o wyjątkowości ludzkiej mądrości, która jest czymś więcej niż tylko prostym rozumowaniem, możliwym do zapisania w algorytmach²³. Z kolei Miguel Benasayag bardzo mocno

²⁰ M. Wojewoda, *Artificial Intelligence as a Social Utopia*, „Ethos” 2023, t. 36, nr 143(3), s. 112, 119–122.

²¹ Tamże, s. 116–118.

²² Franciszek, *Sztuczna inteligencja i pokój*, 1.01.2024, Orędzie na 57. Światowy Dzień Pokoju, <https://www.vatican.va/content/francesco/pl/messages/peace/documents/20231208-messaggio-57giornatamondiale-pace2024.html> (20.01.2024).

²³ L. Korporowicz, *Sztucznej mądrości nie będzie*, „Perspektywy Kultury” 2023, nr 42(3), s. 117–130.

podkreśla znaczenie ludzkiej cielesności, która sprawia, że nasze poznanie, myślenie i działanie nie może być w pełni „odtworzone” przez roboty, wyposażone nawet w super-AI. Wskazuje na wspaniałą zintegrowany system ludzkiego umysłu i cielesności, którego nie da się wyrazić nawet najlepszymi programami i konstrukcjami. Jest to jednak według niego pułapka nowoczesnego racjonalizmu, która może zagrozić człowiekowi²⁴. Wiąże się to z teorią „poznania ucieleśnionego”, rozwijaną wspólnie przez wielu myślicieli. Wskazuje ona na poszerzony zakres tego, co człowiek może „odebrać” ze świata, poznać, nie tylko przez swój rozum, ale też całą swoją cielesność, a czego zarazem nie jest w stanie przekazać robotowi²⁵. Z tych względów próbuje się również wykorzystać ludzkie bądź zwierzęce ciało jako system wykonawczy dla AI.

Człowiek, który ma polecenie, aby coś wykonać, może tego nie zrobić nie tylko z powodu lenistwa, które czasem okazuje się błogosławieństwem, ale również kierując się pewnym przecuciem, którego maszyny mieć nie mogą. Czasami może to być bardzo ważne dla ocalenia jakiegoś dobra czy nawet ludzkiego życia. Na ile jesteś w stanie nauczyć AI żartów czy rozpoznawania sarkazmu lub cynizmu? Choć np. chatboty najnowszej generacji mają już według inżynierów częściowo rozpoznawać ton głosu, to wydaje się, że osiągnięcie „ludzkiego poziomu” w tym aspekcie jest również niezbyt realne. Wyrażenie całej złożoności bytu ludzkiego wyłącznie przez ciągi algorytmów jest nie do zrealizowania i ważna jest świadomość tego przy rozwoju tej technologii, a nade wszystko w perspektywie wyposażania jej w pewne narzędzia wykonawcze czy powierzania „odpowiedzialności” za życie ludzi²⁶. Rzeczywistość jest o wiele bogatsza niż liczby i ciągi danych, a przecież już w starożytności pojawiały się podobne pomysły, opierające się na odkryciu harmonii w stworzeniu, jak w przypadku Pitagorasa i jego następców, dla których wszystko było liczbą.

Choć niektóre koncepcje, jak materializm marksistowski, ograniczają człowieka tylko do wymiaru fizycznego, to przecież większość nurtów

²⁴ M. Benasayag, *The tyranny of algorithms*, transl. S. Rendall, Europa, London 2021, s. 15–52.

²⁵ Zob. A. Dąbrowski, *Ucieleśnione poznanie – założenia, tezy i wyzwania*, „Argument” 2021, t. 11, nr 1, s. 13–32.

²⁶ Franciszek, *Sztuczna inteligencja i pokój*, art. cyt; tegoż, *Sztuczna inteligencja i mądrość serca*, art. cyt.

filozoficznych wskazuje również na szczególny jego wymiar duchowy, który w jeszcze większym stopniu wymyka się kwantyfikacji. Bez ducha nie ma wolności człowieka, wolności żadnego bytu. Nie jest to coś, co człowiek może przekazać najdoskonalszym nawet maszynom. Jeśli natomiast powstanie hybryda człowiek-maszyna, która będzie miała wolność działania, jak niebezpieczne może się to okazać dla człowieka?

Co można i co należy czynić?

Rozwój technologii AI w wielu aspektach jest nie do zatrzymania. Niektórzy postulują choćby czasowe wstrzymanie badań i rozwoju AI, aby ludzkość lepiej na ten rozwój przygotować, co może się wydawać bardziej realne. Takie postulaty podnosił już wcześniej m.in. Elon Musk i mimo że sam ponownie zaangażował się w rozwój projektów AI, to nadal ten postulat podtrzymuje²⁷.

Jeśli powstrzymanie całego rozwoju AI okaże się niemożliwe, to bardziej prawdopodobne i realne będzie uzgodnienie powstrzymania się przynajmniej czasowego przed wypuszczeniem do działania autonomicznych jednostek, zwłaszcza z rozwiniętym modułem wykonawczym. Potrzeba dogłębnego sprawdzenia działania technologii AI, jej możliwości i zagrożeń najpierw w poszczególnych działaniach – zadaniach, a w przypadku AI ogólnej – „na sucho”, tzn. bez podłączania systemu wykonawczego. AI będzie realizowała wszystko, na co ma przyzwolenie i czego reguły wprost jej nie zablokują. Trzeba pamiętać ponadto o możliwości potencjalnej ingerencji w daną jednostkę (również przez „karmienie” jej określoną wiedzą) oraz dążenia do jej przeprogramowania.

Najbardziej realna staje się wszelako praca nad odpowiednim ukierunkowaniem rozwoju i wykorzystania AI. Wydaje się, że stosunkowo łatwo będzie odpowiednio zaprogramować jej systemy zadaniowe, które mogą być wykorzystywane w bardzo wielu sferach życia człowieka: w nauce, produkcji, handlu, usługach, leczeniu, edukacji, pomocy społecznej, ochronie

²⁷ J. Narayan, K. Hu, M. Coulter, S. Mukherjee, *Elon Musk and others urge AI pause, citing 'risks to society'*, 29.03.2023, <https://www.reuters.com/technology/musk-experts-urge-pause-training-ai-systems-that-can-outperform-gpt-4-2023-03-29/> (8.02.2024).

bezpieczeństwa, kulturze, rozrywce itd. Nie znaczy to, że nie nastąpią w tym wypadku przy wykorzystaniu AI problemy etyczne, jak chociażby już powstające kwestie prawa własności intelektualnej, nieetycznego wykorzystania w biznesie, w komunikacji, w walce politycznej itd. Istniejące przepisy prawne dotyczące AI traktują głównie o tym, jak należy właściwie ją wykorzystywać i jakich zasad przestrzegać przy jej rozwijaniu.

Sprawa jest bardziej skomplikowana w przypadku wyposażonego w AI bytu o charakterze ogólnym, samouczącym się, autonomicznym i do tego samoreprodukującym się. W przypadku osób ludzkich mamy kilka podstawowych punktów odniesienia dla podejmowanych działań: obowiązujące na danym terenie prawo, wartości i normy kulturowe, wartości i normy religijno-moralne oraz sumienie. Na ile jesteśmy w stanie zaprogramować maszynę tak, aby uwzględnić nie tylko wszystkie reguły, ale też całą skomplikowaną konstelację tych reguł i ich powiązań? Trzeba znowu jasno powiedzieć: nie jest to w pełni możliwe. Co zatem należy robić i czego możemy maszyny nauczyć?

Najłatwiejsze wydaje się prawo. Trzeba jednak pamiętać o jego ograniczeniu – nie wszystko jest i może nim być regulowane, należy też pamiętać o trudnościach w uchwalaniu prawa, o jego różnicach w różnych krajach, różnicach w jego interpretacji, o prawnych „sztuczkiach”, w których AI zapewne również szybko prześcignie człowieka. Będzie „znała” dokładnie różne systemy prawne i ich słabości. Czy ludzie będą w stanie ujednoczyć prawo i je doprecyzować, dokonywać jednolitej interpretacji? W najbliższej przyszłości na pewno nie, a rozwój AI postępuje bardzo szybko. Nawet jeśli w prawnym regulowaniu rozwoju AI uda się dojść do porozumienia na poziomie globalnym²⁸, to nie jesteśmy w stanie wszystkiego przewidzieć i prawnie zabezpieczyć. Aczkolwiek jest to jedno z działań, które należy podejmować i realizować, tworząc możliwie szybko przejrzyste, kompletne i jednolite przepisy, które wraz z rozwojem technologii i różnymi zdarzeniami wynikającymi z ich zastosowania będą wymagały również nieustannego aktualizowania. Jednakże i tu pojawia się kolejna wątpliwość: co z doskonaleniem i stałą aktualizacją w przypadku jednostek autonomicznych?

²⁸ Franciszek, *Sztuczna inteligencja i mądrość serca*, art. cyt.

Nie mniejsze różnice i trudności są związane z wartościami i normami kulturowymi. Trudno spotkać w literaturze rozważania na ich temat w tym kontekście. W życiu społeczeństw mają one jednak istotne znaczenie i zupełne ich pomijanie w przypadku tej technologii wydaje się ważnym brakiem. Łatwo je uznać po prostu za względne i przejściowe, ale ukazują one sposób życia człowieka tu i teraz, dając mu poczucie bezpieczeństwa w jego działaniach. Jeśli są nieprzestrzegane, dokonuje się degradacja ładu społecznego i następuje zagrożenie dla funkcjonowania człowieka. Z drugiej strony nauczenie tych reguł systemu AI wydaje się nawet łatwiejsze, zwłaszcza że mniejsze niedopasowania i niezgodności są w tym wypadku w mniejszym stopniu obciążone negatywnymi konsekwencjami. Dla jednostek z AI poznanie wszystkich systemów kulturowych i umiejętność dostosowania się do nich nie powinna stanowić większej trudności. Staje się to zarazem kolejnym elementem ukierunkującym ich działania i zabezpieczającym przed naruszeniem dobra człowieka.

Rozważania dotyczące ukierunkowania działań AI koncentrują się najczęściej na kwestii „nauczenia” jej moralności, czy raczej wprowadzenia danych do działania zgodnie z zasadami moralnymi²⁹. Rozważań na ten temat jest już w literaturze dużo i skupiają się one na tym, w jaki sposób zagwarantować to, żeby owe zasady wpisane sztucznym jednostkom były realizowane właściwie. W nawiązaniu do wcześniejszych zagadnień warto jednak zadać pytanie, jakich zasad moralnych mamy je „uczyć”, istnieją przecież także różne systemy moralne, i do tego różnice w ich obrębie. Różnice dotyczą też tego, co rozumie się przez dobro człowieka. Można mówić o pewnych podstawowych normach, co do których panuje powszechna zgodność, ale takich norm jest stosunkowo niewiele. Często systemy moralne są powiązane z systemami religijnymi. To, co jest proponowane przez różnych autorów, to pewne wybrane zasady, które nie są w stanie wyczerpująco wyznaczyć sposobów działania w bardzo różnych sytuacjach, w których nawet człowiek przeżywa rozterki moralne. Ponadto kłócąc się o szczegóły w naszych ludzkich sporach, czasem zapominamy o podstawowych zasadach, które nawet trudno nam sformułować, bo „są oczywiste”, jak: dążenie do dobra, szacunek, wolność, godność ludzka i godność każdego stworzenia itd.

²⁹ Można tu wymienić np. M. Chojnowski, *Etyka sztucznej inteligencji*, Centrum Etyki Technologii Instytutu Humanites, Warszawa 2022.

W przypadku człowieka, gdy brak odpowiednich norm zewnętrznych, ostatecznym „ratunkiem” może być sumienie. Wielu myślicieli mówi o „sumieniu” również w odniesieniu do maszyn, które można zaprogramować, jednak ponownie jest to kwestia wyłącznie możliwości odpowiedniego wgrania przepisów prawa i zasad moralnych. Nic więcej. Nie może tu być mowy o osobistym rozeznawaniu, dialogu wewnętrznym, formowaniu się własnej decyzji moralnej, a tym bardziej odwołania do prawa naturalnego odczytywanego w sumieniu w zetknięciu z określonym dylematem moralnym. Krzysztof Wieczorek, analizując zagadnienie „sumienia” maszyn, rozważa, jak nauczyć maszyny etycznej inteligencji. Według niego jest to działanie podstawowe, możliwe do zrealizowania i najbardziej skuteczne, gdyż stanowi rozwiązanie problemów u samego początku. Uważa on, że trzeba się zastanowić nad sposobami wdrożenia „strategii[i] projektowania procesów uczenia się przez wzmacnianie w taki sposób, aby na każdym etapie procesu wybory, decyzje i działania AI, które sprzyjają człowiekowi i jego żywotnym interesom, otrzymały pozytywne wzmocnienie”³⁰. Ma to ją uczyć działania zawsze z uwzględnieniem dobra człowieka, za co będzie „nagradzana”. Czy jednak jesteśmy w stanie zaprojektować technologię w ten sposób, bez ryzyka błędu, opuszczenia czegoś? A nade wszystko – nawiązując do wcześniejszych rozważań – czy jesteśmy w stanie uwzględnić całą złożoność człowieka, jego życia i moralności, wyjątkowość jego poznania w sumieniu? Podjęcie prób możliwie kompletnego wyuczenia maszyn reguł ich moralnego działania jest jednak bez wątpienia również jednym z elementów działań na rzecz właściwego wykorzystania AI.

Należy niemniej pamiętać, że w wypadku systemów AI, nawet najbardziej doskonałych, nie można mówić o moralnej odpowiedzialności, gdyż mimo wszystko maszyna nie może stanowić podmiotu moralnego. Można jedynie uznać, że w pewnym momencie stanowi zagrożenie i należy ją wycofać z działania. Kto więc jest odpowiedzialny? Czy tylko twórca danej jednostki? Owszem, może on być odpowiedzialny, jeśli zaniedbał coś świadomie lub wprowadził algorytmy, które umożliwiają niewłaściwe działania. Tak jak wielokrotnie już było stwierdzone, nikt nie jest w stanie

³⁰ K. Wieczorek, *The Conscience of a Machine? Artificial Intelligence and the Problem of Moral Responsibility*, „ER(R)GO. Theory–Literature–Culture” 2021, nr 42(1), s. 22.

przewidzieć wszystkiego ani zaprogramować maszyny w taki sposób, aby uwzględnić całą złożoną rzeczywistość, która nas otacza. Nikt też nie jest w stanie przewidzieć, czego się taka jednostka nauczy sama. Odpowiedzialność spoczywa także na osobach stanowiących prawo, na politykach, na osobach decydujących o dopuszczeniu danych jednostek do wykorzystania, zwłaszcza jeśli będą to jednostki autonomiczne. W następnej kolejności jest to odpowiedzialność osób tworzących system prewencyjny, zabezpieczenia, system bezpieczeństwa. Ostatecznie jest to odpowiedzialność każdej osoby współtworzącej „środowisko informacyjne”, którym „karmić się” będzie AI (np. treści namawiające do agresji, przemocy, odrzucenia norm moralnych). Można zatem wyróżnić przynajmniej kilka grup odpowiedzialności moralnej, co winno mieć wyraz w prawie, choć nie wszystko będzie można połączyć bezpośrednio z funkcjonowaniem danej jednostki wyposażonej w AI.

Poza samymi systemami normatywnymi kluczowe znaczenie mają odpowiednio ukształtowane systemy zarządzania społecznego, które niestety będą najpewniej również wspomagane przez AI³¹. Potrzebny jest w tym wypadku system nadzorujący, w którym decyzje zależą tylko od człowieka. Łączy się to ze stworzeniem odpowiednich systemów bezpieczeństwa dla całych społeczeństw, ale również w stosunku do każdej autonomicznej jednostki wyposażonej w AI. Systemy te powinny mieć przynajmniej podwójny układ zabezpieczeń i możliwości wyeliminowania takiej jednostki. System bezpieczeństwa powinien wymagać ponadto stworzenia rozwiązań instytucjonalnych, tzn. powołania odpowiednich urzędów nadzoru czy kontroli nad rozwojem technologii, ale też nad funkcjonowaniem zwłaszcza samodzielnych systemów.

Należy uwzględnić również od samego początku projektowania tych systemów kwestie chęci niewłaściwego ich wykorzystania, ponownie: zwłaszcza systemów autonomicznych, do niegodziwych celów, w tym również wojen czy celów terrorystycznych. Prace nad szerszym wykorzystaniem AI w prowadzeniu działań wojskowych są już rozwijane. Na ile można je powstrzymać? Uzasadnienia wydają się przekonujące: oszczędzić życie ludzi

³¹ Por. M. Wojewoda, art. cyt.

(walczą maszyny) i zapobiec uzyskaniu przewagi przez systemy niedemokratyczne, reżimy lub grupy terrorystyczne. Już w prowadzonej obecnie wojnie na Ukrainie czy w działaniach wojskowych w strefie Gazy systemy z AI są wykorzystywane, choć na razie mają one charakter zadaniowy³². Zagrożenie związane ze stosowaniem systemów autonomicznych będzie wzrastać w miarę zaawansowania AI, ale również wraz z rozwojem systemu wykonawczego; autonomiczne drony, choć mogą mieć szerokie zastosowanie, także na wojnie, nie stanowią jeszcze zagrożenia dla całej ludzkości. Na pewno jednak trzeba zapobiegać możliwości przekierowania systemów autonomicznych (które będą też coraz bardziej zaawansowane), stworzonych w celach pokojowych, na zadania militarne.

Podjęcie różnych działań na rzecz szeroko rozumianej edukacji, z zadaniem poznania tego, czym AI jest, a czym nie jest, jak może być wykorzystywana i jakie zagrożenia się z nią wiążą, jest w najbliższej perspektywie jednym z kluczowych zadań. Niedawno tego typu działanie zostało podjęte przez Ministerstwo Edukacji i Nauki, które na Zintegrowanej Platformie Edukacyjnej zamieściło poradnik dla nauczycieli³³. Jest to jeden z elementów działań, które należy podejmować. Niezbędne jest również uświadamianie i edukacja ludzi przez media, zarówno tradycyjne, jak i społecznościowe, wykorzystanie w tym celu różnych instytucji publicznych czy prowadzenie akcji społecznych w terenie.

Działania edukacyjne w zakresie AI są istotnym elementem prewencji jej niewłaściwego wykorzystywania³⁴. Działaniem roztropnym, także z perspektywy edukacyjnej, wydaje się zadecydowanie od początku rozwoju tej technologii o rezygnacji z prób tworzenia bliskiego podobieństwa zewnętrznego robotów humanoidalnych do ludzi. Pozwoli to uniknąć przynajmniej niektórych zagrożeń, a zarazem osoby edukowane w temacie sposobów postępowania z takimi systemami nie będą obciążone ryzykiem pomyłki.

³² S. Feldstein, *AI in war: Can advanced military technologies be tamed before it's too late?*, 11 January 2024, <https://thebulletin.org/2024/01/ai-in-war-can-advanced-military-technologies-be-tamed-before-its-too-late/> (9.02.2024).

³³ Ministerstwo Nauki, *Do czego AI nie służy. Przewodnik dla nauczycieli*, Ministerstwo Edukacji i Nauki, Zintegrowana Platforma Edukacyjna, <https://zpe.gov.pl/ai-przewodnik-dla-nauczycieli> (3.02.2024).

³⁴ Franciszek, *Sztuczna inteligencja i mądrość serca*, art. cyt.

Zasada większego bezpieczeństwa moralnego wskazuje również na potrzebę rezygnacji z tworzenia hybryd maszyna-człowiek czy nawet zwierzę-maszyna. Stanowi to bowiem poważne naruszenie godności człowieka, a także wartości innych żywych czujących stworzeń. Dopuszczalne natomiast jest takie ingerowanie w organizmy żywe z wykorzystaniem technologii, w tym AI, które nie powoduje degradacji ich wartości, a służyć ma jedynie wsparciu człowieka czy zwierzęcia w jego funkcjonowaniu, uzupełnieniu jego defektów i braków.

W tych dwóch wypadkach występuje największe zagrożenie stworzenia jednostek: urządzenia czy hybrydy, które zamiast być pomocnymi w życiu człowieka i rozwoju społecznym, mogą stać się jego rywalem, dążącym do eliminacji przeciwnika. Jest to zasada działania obecna w życiu społeczeństw niestety od niepamiętnych czasów, i AI na pewno nie może o tym nie wiedzieć.

Zakończenie

Papież Franciszek w Orędziu na Światowy Dzień Środków Społecznego Przekazu przywołał ważne słowa Romano Guardiniego: „Naszym zadaniem jest twórczo wpływać na proces zmian (...), zachowując uczciwą wrażliwość na wszystko, co jest w nim niszczycielskie, nieludzkie”³⁵. Rozwijana technologia AI ma olbrzymi potencjał i, jak się wydaje, nie można tego rozwoju już zatrzymać, nawet czasowo, ale należy czynić wszystko, aby zapobiec temu, co może stanowić zagrożenie dla poszczególnego człowieka, jak i dla całej ludzkości. Uruchomianie kolejnych chatbotów komunikujących się ze zbiorowością globalną i samouczących się wprowadziło nie tylko gwałtowny wzrost zainteresowania tą technologią, zwiększenie świadomości jej możliwości, ale także proces szybkiego samodoskonalenia się systemów AI.

Rozwój ten u niektórych rodzi euforię i zachęca do prezentowania świetlanej przyszłości dzięki AI, ale wśród wielu rodzi obawy, a nawet prowokuje do skrajnie pesymistycznych wizji zagłady ludzkości. Wszystko zależy od tego, na ile człowiek nad tym rozwojem zapanuje, tak jak miał zapanować

³⁵ R. Guardini, *Listy nad jeziora Como*, tłum. K. Markiewicz, Narodowy Instytut Architektury i Urbanistyki, Warszawa 2021, s. 82.

nad całą ziemią (Rdz 1,28), a nie stworzy systemy, które mogą stać się zagrożeniem dla niego samego, rywalem, który będzie chciał wyeliminować swojego przeciwnika.

Potrzebna jest w tym solidarna współpraca krajów i ośrodków badawczych, ich porozumienie dla dobra ludzkości. Wymaga to również jak najszerszych, kompleksowych działań: połączenia elementów polityki i zarządzania, systemów bezpieczeństwa, regulacji prawnych, kształtowania zasad etycznych, edukacji, odpowiedniej komunikacji. Ponadto wymienione aspekty społeczne i etyczne winny być nieustannie na bieżąco konfrontowane ze zmianami technologicznymi. Powinna być dokonywana ocena moralna kwestii związanych z aktualnym etapem rozwoju technologicznego i sposobem wykorzystywania technologii, stanowiący zaś prawo oraz twórcy technologii i odpowiedzialni za ich wykorzystanie winni brać pod uwagę podnoszone przez etyków i socjologów potencjalne problemy, aby unikać od początku pułapek niewłaściwego rozwoju.

Pozytywnie nastawieni do zmian realiści wskazują, że mimo potencjalnych zagrożeń i małych szans na uniknięcie wielu z nich rozwój technologii AI i jednostek w nią wyposażonych pozwoli odkrywać pełniej godność i wyjątkowość człowieka oraz wartość świata stworzonego, w którym żyjemy. Paradoksalnie zatem tworząc „konkurencję” – konkurenta dla człowieka, będziemy musieli odkryć to, co nas wyróżnia, co jest specyficznego, naszą godność i wielkość, niesprowadzalność do najbardziej rozwiniętego bytu z AI. Doceniona zostanie cała integralność bytu ludzkiego, łącznie z jego cielesnością i duchem oraz relacyjnością. Maszyny mogą stanowić pewne substytuty w relacyjności międzysobowej, ale nigdy nie zastąpią w tym drugiego człowieka. Maszyna nie potrzebuje miłości, tylko źródła zasilania; nie jest w stanie obdarzyć miłością, nawet jeśli będzie bardzo użyteczna czy pomocna³⁶. Specyfika bytu ludzkiego opiera się również na niemożności odrzucenia przez niego sensu własnego istnienia. Właśnie pod wpływem konfrontacji z technologiami robotycznymi ujawnia się wartość człowieka i jego godność³⁷.

³⁶ Zob. P. Donati, *A critical realist perspective on humanness as a meaningful re-entry of relational distinctions*, „Journal for the Theory of Social Behaviour” 2021, vol. 51(1), s. 49–71.

³⁷ P. Donati, *Impact of AI/Robotics on Human Relations: Co-evolution Through Hybridisation*, w: *Robotics, AI, and Humanity*, ed. J. von Braun et al., Springer, Cham 2021, s. 213–227.

Bibliografia

- Alichniewicz A., Michałowska M., *Udoskonalanie gatunku ludzkiego. Problemy definicyjne i argumentacyjne*, „Principia” 2018, nr 64, s. 231–261
- Archer M. S. (ed.), *Social Morphogenesis*, Springer, London–New York 2013.
- Benasayag M., *The tyranny of algorithms*, transl. S. Rendall, Europa, London 2021.
- Benedykt XVI, *Caritas in veritate. O integralnym rozwoju ludzkim w miłości i prawdzie*, Watykan 2009.
- Bostrom N., Roache R., *Ethical issues in human enhancement*, w: *New Waves in Applied Ethics*, ed. J. Ryberg, T. Petersen, C. Wolf, Palgrave Macmillan, Basingstoke 2007, s. 120–152.
- Braun J. von, Archer M. S., Reichberg G. M., Sánchez Sorond M., *AI, Robotics, and Humanity: Opportunities, Risks, and Implications for Ethics and Policy*, w: *Robotics, AI, and Humanity*, ed. J. von Braun et al., Springer, Cham 2021, s. 1–13.
- Cebo D., *Human enhancement*, „SSRN Electronic Journal” January 2020, doi: 10.2139/ssrn.4170853.
- Chojnowski M., *Etyka sztucznej inteligencji*, Centrum Etyki Technologii Instytutu Humanites, Warszawa 2022.
- Ciesielski M., *Generatywna AI to dopiero początek. Nadchodzą autonomiczni agenci*, 11.01.2024, <https://forsal.pl/lifestyle/technologie/artykuly/9398795,generatywna-ai-to-dopiero-poczatek-nadchodza-autonomiczni-agenci.html> (7.02.2024).
- Dąbrowski A., *Ucieleśnione poznanie – założenia, tezy i wyzwania*, „Argument” 2021, t. 11, nr 1, s. 13–32.
- Demartini G., D. E. Difallah D. E., Gadiraju U., Catasta M., *An Introduction to Hybrid Human-Machine Information Systems*, „Foundations and Trends in Web Science” 2016, vol. 7, no. 1, s. 1–87.
- Demartini G., Difallah D. E., Gadiraju U., Catasta M., *An Introduction to Hybrid Human-Machine Information Systems*, Now Foundations and Trends, Boston 2017.
- Donati P., *A critical realist perspective on humanness as a meaningful re-entry of relational distinctions*, „Journal for the Theory of Social Behaviour” 2021, vol. 51(1), s. 49–71.
- Donati P., *Impact of AI/Robotics on Human Relations: Co-evolution Through Hybridisation*, w: *Robotics, AI, and Humanity*, ed. J. von Braun et al., Springer, Cham 2021, s. 213–227.
- Feenberg A., *Democratizing Technology: Interests, Codes, Rights*, „The Journal of Ethics” June 2001, vol. 5(2), s. 177–195.
- Feldstein S., *AI in war: Can advanced military technologies be tamed before it's too late?*, 11 January 2024, <https://thebulletin.org/2024/01/ai-in-war-can-advanced-military-technologies-be-tamed-before-its-too-late/> (9.02.2024).
- Feuerriegel S., Hartmann J., Janiesch Ch., Zschech P., *Generative AI*, „Business & Information Systems Engineering” 2024, 66, s. 111–126, online: wrzesień 2023, doi: [org/10.1007/s12599-023-00834-7](https://doi.org/10.1007/s12599-023-00834-7).
- Ferdynus M., *Czy mamy prawo „oswoić” sztuczną inteligencję*, „Ethos” 2023, t. 36, nr 143(3), s. 67–91.

- Franciszek, *Sztuczna inteligencja i pokój*, 1.01.2024, Orędzie na 57. Światowy Dzień Pokoju, <https://www.vatican.va/content/francesco/pl/messages/peace/documents/20231208-messaggio-57giornatamondiale-pace2024.html> (20.01.2024).
- Franciszek, *Sztuczna inteligencja i mądrość serca: dla komunikacji w pełni ludzkiej*, 24.01.2024, Orędzie na 58. Światowy Dzień Środków Społecznego Przekazu 2024, <https://www.vatican.va/content/francesco/pl/messages/communications/documents/20240124-messaggio-comunicazioni-sociali.html> (5.02.2024).
- Gartner Research, *Predicts 2019: The Democratization of AI*, 29 November 2018, <https://www.gartner.com/en/documents/3894386> (7.02.2024).
- Gartner Research, *Generative AI Can Democratize Access to Knowledge and Skills*, October 17, 2023, <https://www.gartner.com/en/articles/generative-ai-can-democratize-access-to-knowledge-and-skills> (23.02.2024).
- Guardini R., *Listy znad jeziora Como*, tłum. K. Markiewicz, Narodowy Instytut Architektury i Urbanistyki, Warszawa 2021.
- Jaba S., *World's First Animal-Machine Hybrids Start to Reproduce*, 17 December 2021, <https://chapterzmagazine.com/2021/12/worlds-first-animal-machine-hybrids-start-to-reproduce/> (8.02.2024).
- Korporowicz L., *Sztucznej mądrości nie będzie*, „Perspektywy Kultury” 2023, nr 42(3), s. 117–130.
- Ministerstwo Nauki, *Do czego AI nie służy. Przewodnik dla nauczycieli*, Ministerstwo Edukacji i Nauki, Zintegrowana Platforma Edukacyjna, <https://zpe.gov.pl/ai-przewodnik-dla-nauczycieli> (3.02.2024).
- Narayan J., Hu K., Coulter M., Mukherjee S., *Elon Musk and others urge AI pause, citing 'risks to society'*, 29.03.2023, <https://www.reuters.com/technology/musk-experts-urge-pause-training-ai-systems-that-can-outperform-gpt-4-2023-03-29/> (8.02.2024).
- Porpora D. V., *Morphogenesis and Social Change*, w: M. S. Archer (ed.), *Social Morphogenesis*, Springer, London–New York 2013, s. 25–37.
- Różanowski K., *Sztuczna inteligencja: rozwój, szanse i zagrożenia*, „Zeszyty Naukowe” 2007, s. 109–135.
- Russell S. J., Norvig P., *Artificial Intelligence. A Modern Approach*, Pearson, New Jersey 2010.
- Searle J. R., *Minds, brains, and programs*, „The Behavioral and Brain Sciences” 1980, vol. 3, s. 417–457.
- Smołucha D., *Cyborg – mariaż człowieka i technologii*, „Perspektywy Kultury” 2023, nr 42(3), s. 191–199.
- Sztompka P., *Socjologia zmian społecznych*, tłum. J. Konieczny, Znak, Kraków 2010.
- Taylor Ch., *'Sapiens' author says AI is an alien threat that could wipe us out: 'Instead of coming from outer space, it's coming from California'*, 12.09.2023, <https://fortune.com/2023/09/12/sapiens-author-yuval-noah-harari-ai-alien-threat-wipe-out-humanity-elon-musk-steve-wozniak-risk-cogx-festival/> (5.02.2024).
- Thang Vo-Doan T., Sato H., *Insect-machine Hybrid System: Remote Radio Control of a Freely Flying Beetle (*Mercynorrhina torquata*)*, „Journal of Visualized Experiments” September 2016, DOI: 10.3791/54260.
- Wieczorek K., *The Conscience of a Machine? Artificial Intelligence and the Problem of Moral Responsibility*, „Theory–Literature–Culture” 2021, nr 42(1), s. 15–34.

Wielecki K., *Kultura versus kultura masowa*, NCK, Warszawa 2024.

Wojewoda M., *Artificial Intelligence as a Social Utopia*, „Ethos” 2023, t. 36, nr 143(3), s. 109–126.

Artificial intelligence – a tool or a rival to humans?

Summary

The text considers current and upcoming social changes that may occur under the influence of the development and use of artificial intelligence technologies. After analyzing the possibilities of different directions for the development of artificial intelligence, the author addresses the issue of maintaining appropriate principles and rules of this development so that they are not destructive to humans and social order. It points to the trap of hasty and unprepared acceleration, which may undermine the stable functioning of societies. Then, an integral system of activities was considered, involving public institutions, research centers, business entities and every conscious user of virtual space to take comprehensive actions to properly direct the development and use of artificial intelligence.

Keywords: artificial intelligence, social development, social change, technology ethics

Sztuczna inteligencja wobec wartości autotelicznych. Pan Twardowski współcześnie

Wstęp: Odwieczne pragnienie transgresji

Człowiek od początku swego istnienia nie tylko pragnął realnych działań na rzecz wielorako pojętej transgresji, lecz także się nimi wykazywał, a więc przekraczał granice tego, „co się ma”, i tego, „kim się jest”. Bez tego pragnienia prawdopodobnie nie zdołałby zmieniać swojego środowiska, narzędzi, którymi dysponował, ale i swojej wiedzy, a w końcu swojej osobowości. Każde dokonane przez niego odkrycie poszerzało obszar jego działania i budowało imperatyw transformacji, który pozostawiano kolejnym pokoleniom niezależnie od tego, ile czyniono dobra, ale i zła i na ile usprawniano procesy odkrywania i budowania jego wolności. Cena, jaką człowiek płacił za możliwość dokonywania transgresji, nie była w istocie poddawana szczególnym rodzajom kalkulacji, nawet wówczas, gdy, jak w przypadku odkrywania broni masowego rażenia, prowadzi to do możliwości uśmiercania nieokreślonej liczby ludzi. Ta uniwersalna dla człowieka prawda w sposób bardzo celny znajduje wyraz w literackim wątku pana Twardowskiego. W roku 1837 znany polski pisarz, historyk, folklorysta i edytor, Kazimierz Władysław

¹ Dr hab. Leszek Korporowicz, prof. Uniwersytetu Jagiellońskiego w Krakowie.

Wójcicki w książce zatytułowanej *Klechdy. Starożytne podania i powieści ludu polskiego i Rusi* pisał:

W starej księdze (Twardowski) raz wyczytał, jak diabła przywołać można: o północnej godzinie cicho wychodzi z Krakowa, kędy leczył w całym mieście, i przybywszy na Podgórze zaczął biesa głośno wzywać. Stanął przeto zawezwany: jak w one czasy bywało, zawarli ze sobą umowę. Na kolanach zaraz diabeł napisał długi cyrograf, który własną krwią Twardowski, wyciśniętą z serdecznego palca, podpisał. (...) Wszystko, co jeno zażądał żywnie, miał na swoje zawołanie: jeździł na malowanym koniu, latał w powietrzu bez skrzydeł, w daleką drogę wsiadał na koguta i prędzej biegał niż konno².

Czyż człowiek współczesny nie chce „żądać żywnie”, aby na jego zawołanie „jeździć na malowanym koniu, latać w powietrzu bez skrzydeł, w daleką drogę wsiadać na koguta i prędzej biegać niż konno? Wszystko wskazuje na to, iż chce on czegoś znacznie więcej, a jego oczekiwaniom nie ma już końca. Chce być silniejszy, podróżować w kosmosie, pokonywać bariery czasu i przestrzeni, co więcej, chce być prawie nieśmiertelny. Z łatwością rezygnuje na rzecz tego pragnienia ze swojego człowieczeństwa, marząc o tym, aby stać się cyborgiem nowej rzeczywistości, postczłowiekiem doby transhumanizmu. Jednym ze środków do tego świata marzeń ma być sztuczna inteligencja. Co dziwne, akceptuje fakt, iż stanie się ona „sztuczna”, aby tylko pokonywała granice tej „naturalnej” i pozwalała na przekroczenie wszelkich innych granic współczesnego pana Twardowskiego. Czy możemy jednak być pewni, że nie będzie on musiał, podobnie jak ten historyczny, salwować się ucieczką, ratując się podróżą na samotny i bezludny Księżyc przyszłości?

² K. W. Wójcicki, *Klechdy. Starożytne podania i powieści ludu polskiego i Rusi*, wybór i opracowanie: R. Wojciechowski, Dragon, Bielsko-Biała 2021.

O przedmiocie i granicach sztucznej inteligencji. Paląca potrzeba zdefiniowania problemu

Naiwność pana Twardowskiego, podobnie jak jego pragnień, nie zna swoich granic. Niestety dziedziczą ją wszyscy jego potomkowie doby cyberrewolucji, przypisując inteligencji cechy i władzę, której nie posiada, tak jak część nigdy nie stanie się całością, tak jak początek nie stanie się końcem, a kwadrat nie stanie się kołem. Fascynaci sztucznej inteligencji zapominają na ogół o prostym fakcie, że inteligencja, nawet w najdoskonalszej swojej wersji, w tym także inteligencja naturalna, pozostaje w dalszym ciągu jedynie inteligencją. Ograniczenie to niesie ze sobą fundamentalne, choć pomijane konsekwencje antropologiczne:

Po pierwsze, inteligencja – bazując na atrybutach inteligencji naturalnej – jest tylko fragmentem szeroko pojętej osobowości. W swej istocie nie wyczerpuje ona całej jej struktury, niezależnie od przyjętej w tym celu teorii i różnic w psychologicznych koncepcjach człowieka. Choć pozostaje ona trwale powiązana ze sferą ludzkich afektów, a więc sferą emocjonalną, sensoryczną, wolicjonalną oraz zdolnościami do przeżywania i projektowania bardzo różnorodnych typów wartości, to nie wyczerpuje nawet poznawczych umiejętności człowieka, odwołujących się do jego kompetencji estetycznych, moralnych i duchowych. Służy ona przede wszystkim rozpoznawaniu i realizacji instrumentalnych potrzeb człowieka, które nie są w stanie rozwiązać jego dylematów etycznych, nie budują jego wrażliwości społecznej, nie definiują jego potrzeb religijnych³.

Po drugie, każdy rodzaj inteligencji oraz sposoby jej wykorzystania, w tym inteligencji sztucznej, uczestniczy tylko we fragmencie procesów socjalizacji i określa tylko fragment procesu i problemu wychowania. Proces socjalizacji, a szczególnie kulturalizacji, polega na szczególnym trybie uczenia się, doświadczania i praktykowania cech kultury symbolicznej konkretnej społeczności. Procesami tymi rządzą reguły, które odwołują się nie tylko do sfery poznawczej człowieka. Budują jego poczucie tożsamości, jaźni

³ J. F. Lyotard, *Kondycja ponowoczesna*, tłum. M. Kowalska, J. Migasiński, Aletheia, Warszawa 1997. Por.: Z. Orbik, *W poszukiwaniu filozoficznych źródeł problematyki sztucznej inteligencji*, „Ethos” 2023, R. 36, nr 3(143), s. 45–66.

społecznej i kulturowej o bardzo silnym komponentcie aksjologicznym oraz interakcyjnym. Cechuje je znaczący poziom empatii. Wymienione przymioty osobowości wybiegają poza obszar inteligencji, decydują, że życie społeczne wymaga czegoś więcej niż sprawności przestępcy, a sztuka rządzenia musi odwoływać się nie tylko do sztuki manipulowania i sterowania ludźmi. W ten właśnie sposób odróżniamy proces socjalizacyjny od wychowawczego, który nie sprowadza się do biernego przyjmowania reguł życia grupowego. Wprowadza on w system wartości i umiejętności krytycznego osądu wiodących celów istnienia i rozwoju wspólnoty. Sama inteligencja tu nie wystarcza i niepodobna wyobrazić sobie sztucznej tożsamości, lojalności oraz patriotyzmu. Cech tych nie wypełni żaden wyrafinowany rodzaj cyborgizacji uczuć społecznych, digitalizacji wierzeń religijnych oraz sztucznej humanizacji stosunków społecznych⁴.

Po trzecie, inteligencja stanowi tylko komponent kompetencji komunikacyjnych i kulturowych człowieka niezbędnych do uczestnictwa w kulturze na wszystkich jej poziomach i obszarach. O kompetencjach tych decyduje splot czynników, których źródłem jest nie tylko funkcja kognitywna ludzkiej inteligencji, dająca człowiekowi zdolność rozpoznawania abstrakcyjnych struktur gramatycznych czy semiotycznych w ogóle w postaci dającego się kodować systemu komunikacyjnego. To ważne odkrycie transformacyjno-generatywnej teorii Noama Chomsky'ego zostało skutecznie uzupełnione przez późniejsze analizy całokształtu kompetencji komunikacyjnych człowieka. O faktycznym ich funkcjonowaniu decydują elementy społecznego kontekstu ich użycia. Ich źródłem nie jest mózg ludzki, tylko praktyka społeczna konkretnej społeczności, tworząca ponadindywidualne i kolektywnie podzielane reguły rzeczywistości kulturowej wraz z ich emocjonalnym sposobem doświadczania, ale także łamania, przewartościowywania i transgresji⁵. Ujawnienie tego typu komponentów kompetencji kulturowej stało się powodem nie tyle krytyki, ile znaczącego poszerzenia wąsko kognitywnych teorii Chomsky'ego i uznania wieloczynnikowej natury powstawania,

⁴ L. Korporowicz, *W poszukiwaniu sztucznego człowieczeństwa*, w: *Rzeczywistość hybrydalna. Perspektywa pedagogiczna*, red. S. Jaskuła, Księgarnia Akademicka, Kraków 2023, s. 79–95.

⁵ R. Wiśniewski, *Transgresja kompetencji międzykulturowych. Studium socjologiczne młodzieży akademickiej*, Wydawnictwo UKSW, Warszawa 2016.

reprodukcji, ale i zmiany umiejętności uczestniczenia w kulturze, które są tyleż faktem psychologicznym, co społecznym. Nie można wyjaśnić ich prawami ludzkiej inteligencji, nawet w przypadku daleko posuniętej atomizacji i psychoneurologicznej redukcji. Wszelkie czynności schematycznego „mapowania” kontekstu kulturowego i swoistego aneksowania reguł kompetencyjnych do konkretnie wskazanych sytuacji nie nadążają za ich zmiennością oraz wkomponowaniem w spontaniczne procesy wartościowania, które dzieją się w sferze logotwórczych dynamizmów kultury⁶.

Po czwarte, inteligencja rodzi tylko fragment wyzwań rozwojowych człowieka, których źródłem są potencjały osobowości i osoby ludzkiej i które wytwarzają wartości odmienne od instrumentalnych. Doświadczenia rozwojowe nie muszą kierować się logiką kumulacji, racjonalnego wyboru, unikania zagrożenia i kalkulacji poniesionych wysiłków. Dlatego przekraczają one inne kryteria doskonalenia ludzkich działań, którym patronują strategie podporządkowane wyborom inteligentnym, w których ceni się poszukiwanie skutecznego sposobu osiągnięcia celu, a mogą nawet być z owymi kryteriami sprzeczne. W przypadku człowieka i jego człowieczeństwa nie bez znaczenia są przy tym cele spełniające atrybuty wartości autotelicznych, definiowanych często jako istota celów rozwojowych, wykraczających poza kryteria dojrzewania, przystosowania, zdobywania kolejnych szczebli statusu społecznego, nie mówiąc już o pozycjonowaniu w hierarchii stosunków władzy. Wartości autoteliczne są ponadto odnoszone nie tylko do indywidualnej kondycji każdego człowieka z osobna, lecz także do jego życia osobowego, łączącego się z życiem innych ludzi oraz ogarniającymi je wspólnotami. One bowiem mogą stanowić obszar realizacji i dopełniającego spełnienia rozwoju osobowego, bez których straciłby on swoją integralność, a co ważne – również pełnię oczekiwanych owoców.

⁶ L. Korporowicz, *Logotwórcze dynamizmy kultury*, w: tegoż, *Socjologia kulturowa. Kontynuacje i poszukiwania*, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków 2011, s. 15–35.

Istota i atrybuty wartości „autotelicznych”

Istnienie wartości autotelicznych jest prawdziwym wyrzutem sumienia współczesnej humanistyki, coraz bardziej skłonnej do porzucenia zainteresowania swoim właściwym przedmiotem, którym jest nie tyle sam człowiek, ile obecne w nim człowieczeństwo. Ucieczka od tak pojętej misji zaczęła się od radykalnych manifestacji behawioryzmu z lat 20. i 30. ubiegłego wieku, nakazujących koncentrację na ludzkich zachowaniach jako najbardziej wiarygodnym elemencie obserwacji bytu ludzkiego. Problem tylko w tym, że w takim ujęciu nie różni się on niczym od zwierzęcego, a zoologiczne niemal inklinacje antropologii kulturowej szkoły psychokulturalizmu amerykańskiego są do tej pory starannie ukrywane. Fakt, iż nauki społeczne i humanistyczne zawsze zdawały sobie sprawę, że te same zachowania mają nieporównywalne często funkcje, a jeszcze bardziej znaczenia, nie zmienił opresywnych ambicji neopozytywizmu i wykluczenia całej plejady humanistycznie zorientowanych uczonych. Jednym z nich był Florian Znaniecki, stojący na uboczu głównego dyskursu naukowego, aż do lat 60. Rewitalizacja zainteresowania autotelicznymi wartościami kultury symbolicznej nie od razu doprowadziła do głębszych rezultatów. Banalne w swej istocie odkrycie, iż te same znaczenia mogą mieć różne wartości, a sama rewolucja przywracająca godność humanistyce musi wykroczyć poza intensywnie eksplorowaną semiologię, nie od razu uwolniło wrażliwość i antropologiczną świadomość nowego pokolenia badaczy.

Odwołanie się do atrybutów wartości autotelicznych, zagubionych i zatraconych w spuściznie redukcji aksjologicznych współczesnego materializmu, postmodernizmu oraz cyfrowego zamachu na humanizm, może być drogą częściowej przynajmniej rehumanizacji epoki sztucznej inteligencji. Wartości autoteliczne mogą być określane przez zespół następujących cech:

Po pierwsze, nie stanowią one środka do realizacji innych wartości. Są zasadniczo nieinstrumentalne. Ich realizacja nie jest elementem w łańcuchu czynności, których spełnienie służy innym, stojącym poza nimi celom, do których są jedynie zbliżeniem. Stanisław Ossowski, jeden z najwybitniejszych polskich socjologów wartości, a za nim Antonina Kłoskowska, której pokaźna liczba teorii dotyczy wartości autotelicznych w kulturze symbolicznej, dawali przykład wartości estetycznych jako znaczących egzemplifikacji doznań

niedających się sprowadzić do innych i niepodporządkowanych innym ich funkcjom⁷. Ossowski, podkreślając ważną w osobowości człowieka umiejętność refleksyjnego „życia chwilą”, podkreślał, iż w dobie instrumentalnie zorientowanej cywilizacji współczesnej wartości estetyczne stanowią wyjątkową jakość doświadczeń ludzkich o walorach antropologicznych niezależnie od treści i wzorców konkretnie istniejącej kultury⁸. Nie dają się one uczynić narzędziem żadnej psycho- lub socjotechniki, inżynierii społecznej sterowanej zarówno naturalną, jak i tym bardziej sztuczną inteligencją. Ich autotelizm przejawia się także w innych obszarach kultury, w odniesieniu do innego typu doświadczeń w sferze relacji międzyludzkich, uczuć społecznych i religijnych⁹. Z tego powodu wynika ich kolejna cecha:

Po drugie, nie można ich zastąpić innymi wartościami o analogicznym charakterze. Nie mają substytutów i zamienników, które można by użyć lub na które można się powołać w zastępstwie w przypadku ich braku lub wyczerpywania. Takimi wartościami są dla przykładu godność, honor, miłość czy odpowiedzialność. Trudno wyobrazić sobie, aby można było je kupić lub wymienić na odpowiadające im substytuty. Być może dlatego wartości autoteliczne stanowią definiujące atrybuty człowieczeństwa, od których ucieczka stanowi jego pogwałcenie. Na tych wartościach opiera się fundament praw człowieka, praw kulturowych czy prawa narodów, przyznający osobie ludzkiej, ale i wspólnocie prawo do tożsamości, prawo do obrony swej godności czy ciągle nierozumiane i niedoceniane prawo do ojczyzny, patriotyzmu i religii¹⁰. Nie oznacza to akceptacji przymusowej socjalizacji do tych trzech fenomenów człowieka i kultury. Lęk przed takiego rodzaju praktyką w żaden sposób nie może stanowić ideologicznego oręża otwarcie

⁷ A. Kłoskowska, *Socjologia kultury*, PWN, Warszawa 1981.

⁸ S. Ossowski, *U podstaw estetyki*, PWN, Warszawa 1966.

⁹ A. Kłoskowska, *Koncepcja autoteliczności kultury*, w: *O społeczeństwie i teorii społecznej. Księga poświęcona pamięci Stanisława Ossowskiego*, red. M. Ofierska, Z. Mokrzycki, J. Szacki, PWN, Warszawa 1985.

¹⁰ W. B. Gudykunst, R. B. Shapiro, *Communication in everyday interpersonal and intergroup encounters*, „International Journal of Intercultural Research” 1996, 20/1, s. 19–45. Por. także: S. Jaskuła, *Bezpieczeństwo kulturowe w świecie hybrydalnym. Aspekty teoretyczne i metodologiczne*, w: *Granice tożsamości nauk o bezpieczeństwie. Perspektywa materialna i formalna*, red. T. Kośmider, W. Kitler, Difiń, Warszawa 2017, s. 202–218.

zwalczającego kojarzone z nimi postawy w celu budowania światów alternatywnych, o łatwiejszym poziomie sterowania.

Po trzecie, konsekwencją cech poprzednich jest ich finalny, logotwórczy sens pozwalający je określać jako wartości „same w sobie” lub „same dla siebie”, których uzasadnienia i przyczyny ich istnienia nie poszukuje się w przedmiotowej, środowiskowej i zewnętrznej w stosunku do nich wizji człowieka. Dlatego też, w perspektywie wartości autotelicznych, nie może istnieć ani sztuczna mądrość, ani sztuczna godność, sztuczna miłość, sztuczny honor, sztuczna religia. Podobnie nieautentyczne, a więc wewnętrznie puste i nietrwałe, będą takie iluzje jak: sztuczna ojczyzna, sztuczna wspólnota czy sztuczna odpowiedzialność.

Wymienione wyżej atrybuty odpowiadają na potrzebę odnalezienia często zagubionych kryteriów oceny ludzkich transgresji. One same traktowane są mylnie jako wystarczająca, definiująca i najbardziej charakterystyczna emanacja człowieczeństwa. Gdyby stanowisku temu przyznać słuszność, niemal każdy podzieliłby losy pana Twardowskiego, starając się znaleźć sposób na stanie się superczłowiekiem. Sztuczna inteligencja stanowiłaby wystarczającą podstawę do zbudowania świata sztucznego człowieczeństwa o transhumanistycznej tożsamości. Byłby to jednak projekt, na który pan Twardowski wbrew pozorom nie miałby już żadnego wpływu, miotany jedynie ogonem swojego nowego pana i boleśnie uderzając o wszystko, co będzie stanowić o dalece nie jego i nie dla niego zbudowanym świecie. Problem w tym, że cele, które powodują i ukierunkowują przechodzenie granic tego, co człowiek ma, i tego, kim jest, nie mogą jednokierunkowo i wybiórczo rozbudowywać jedynie hubrystycznych i zaborczych mrzonek człowieka. Transgresje mogą bowiem być tyleż wyzwalaające, co zniewalające. Znacznie ważniejsze są potencjały rozwojowe, rozsadzające niejako od środka wielką bańkę nadmuchaną tylko fragmentem ludzkich możliwości. Zboczenie z kursu równoważenia wielorakich konsekwencji transgresji człowieka, dopełniania, ale i harmonizowania jego potrzeb nader często prowadzi w ślepią uliczkę, na której spotyka pułapki swoich nader jednostronnych projekcji. Uliczka ta, pełna postczłowieczych śladów pana Twardowskiego, nie znajduje się już niestety na Ziemi.

Konsekwencje teoretyczne i praktyczne

Warto zapytać, jakie to nauki przynosi nam współczesny pan Twardowski, porzucający koncepcję wartości autotelicznych, opanowujący sztukę optymalizacji swoich działań, cieszący się swą nieśmiertelnością. Nie wiemy jednak, czy miał wyrzuty sumienia oraz czy zdołał je porzucić. Czy oderwał się od swoich tęsknot i wyeliminował dysfunkcjonalne emocje? Nie wiemy, czy odrzucił uprawianie filozofii, a modlitwa nie jest mu już potrzebna. Czy wobec tego w ogóle coś wiemy? Otóż tak, wiemy bardzo wiele, i wiedzę tę możemy podzielić na świadomość teoretycznych i praktycznych dokonań pana Twardowskiego, który powieła swą przygodę w nowej scenarii dobrze znanego mu wszak miejsca. Sztuczna inteligencją karczma ta się nazywa.

W wymiarze teoretycznym wskazać należy co najmniej trzy konsekwencje kulturowej aneksji humanistycznego postrzegania i rozumienia kondycji człowieka współczesnego: Są to:

1. **Marginalizacja pytania o podmiotowość człowieka** w postaci dominacji przywołanych już pytań o to, **czym**, nad to, **kim** jest człowiek. Ewentualna podmiotowość osobowego wymiaru człowieka zamieniona jest na problem **sprawstwa, autonomii, innowacyjności**, a czasem po prostu **produktywności** człowieka przedsiębiorczego, zainteresowanego awansem społecznym przez optymalny wybór wariantu kariery zawodowej w konkretnym środowisku pracy¹¹. Tożsamość osobowa, kulturowa i społeczna zastąpiona zostaje konfiguracją wielorakich, czasami kontekstualnych, na ogół zmiennych identyfikacji, sterowanych wymogami kręgów społecznych z portalami społecznościowymi włącznie. Sztuka gry interakcyjnej stanowi w nich rodzaj inteligencji społecznej na scenie strategii budowania wizerunków i marketingu osobistego w towarzystwie umiejętności oddziaływania, uwodzenia i perswazji.

2. **Ucieczka od atrybutów człowieczeństwa** w koncepcje posthumanizmu, wyzwania projektu *conceptual age*¹², transhumanizmu oraz coraz większe

¹¹ J. Życiński, *Personalizm po śmierci podmiotu ludzkiego*, w: *Personalizm polski*, red. M. Rusecki, Wydawnictwo KUL, Lublin 2008, s. 9–20.

¹² D. Pink, *A whole new mind: Moving from the Information Age to the Conceptual Age*, Riverhead Hardcover, New York 2005.

niezrozumienie i kwestionowanie działań opartych na wartościach autotelicznych. Nacisk na aksjologię optymalizacji oraz eliminowanie prawa do błędu i logiki regresji programowo zamienia wartości na funkcje, treści na informacje, komunikację na przekaz, rozwój na postęp, osobę na jednostkę, wychowanie na socjalizację, wspólnotę na społeczność. Towarzyszące takiemu rodzajowi progresji kompetencje mają być w coraz większym stopniu **cyfrowe**, podobnie jak kulturoznawstwo, czy w ogóle humanistyka, utożsamiając narzędzia i instrumenty poznawania z ich obiektami, celami oraz dającym się obronić sensem ich pozainstrumentalnego istnienia. Za całkiem normalne wielu przedstawicieli nauk społecznych i humanistycznych uznaje dziś przywołane pojęcie kulturoznawstwa cyfrowego, ale z pewnością za niedorzeczne uznaliby pojęcie kulturoznawstwa elektromagnetycznego. Tymczasem technologie cyfrowe są takim samym rodzajem narzędziowego wsparcia nauk o kulturze, jak kiedyś urządzenia elektromagnetyczne w rodzaju magnetofonu, radia czy telewizji. Nikomu nie przychodziło jednak na myśl, aby mitologizować ich użycie, wprowadzając w nazwę dyscypliny podporządkowanej dalece niemagnetycznej metodologii. Ta prosta w swej istocie analogia pokazuje stopień infantylizacji i dehumanizacji wrażliwości naukowej współczesnej posthumanistyki, biegnącej na oślep śladami pana Twardowskiego, uwiedzionej obietnicami „latania w powietrzu bez skrzydeł” i magicznego użycia koguta, w naszym przypadku – sztucznej inteligencji, co „pobieży szybciej niż konno”.

3. Redukcjonizm aksjologiczny. Eliminacja wartości autotelicznych w aksjologicznym horyzoncie sztucznej inteligencji prowadzi do poważnej redukcji kilku fundamentalnych zasad tworzących znane nam oblicze kultury:

- a) Wyeliminowanie kategorii godności niemal bezpośrednio uderza w dotychczasowe rozumienie praw człowieka, któremu przysługuje wszelako, aby nie postrzegać go jedynie przez formalnie określone role, co więcej – aby nie odbierać mu prawa do wspólnoty pojętej jako wartość przekraczająca sumę jednostek oraz ważna część stanowiąca o jego tożsamości¹³. Ona także, poza swoim aspektem instrumentalnym,

¹³ L. Korporowicz, *Polskie korzenie prawa narodów w budowaniu podmiotowości wspólnot kulturowych*, w: *Eseje o niepodległej*, red. R. Wiśniewski, Narodowe Centrum Kultury, Warszawa 2018, s. 102–129.

jest odczuwana jako wartość autoteliczna. Nie daje się więc zamienić na inną, zastąpić wartością analogiczną; nie sposób też wypracować jakiś rodzaj jej podróbki. Zmiana tożsamości jest oczywiście możliwa, ale ponownie musi to być tożsamość o tych samych atrybutach i swoistym miejscu w całości kształcie osobowego życia człowieka¹⁴. Sztuczna tożsamość, podobnie jak sztuczna wspólnota, oznaczać może jedynie dobrze już znaną i krytycznie ocenianą przesocjalizowaną koncepcję człowieka, charakteryzującą autokratyczne i totalitarne systemy społeczne.

- b) Kolejna redukcja aksjologiczna generowana przez instrumentalno-racjonalną dominantę posthumanistycznej wizji człowieka jest ukierunkowana na sprowadzanie autotelicznych aspektów fenomenu ludzkiej mądrości najpierw do wiedzy, a później już tylko do informacji. To właśnie obecność wartości autotelicznych pozwala kategorii te odróżniać, pozwalając tej pierwszej na transgresję, a nawet transcendencję czystego rozumu i władzy sądzenia oraz na odrzucenie umiejętności posługiwania się wiedzą, jeśli jest ona ludobójcza, kryminalna i poniża ludzkie człowieczeństwo. Sztuczna inteligencja jest tu tylko elementem szerzej zakrojonej strategii i nie działa samodzielnie. Z tych też powodów dokonywane za jej pomocą redukcje zakładają odmienny sens posthumanistycznej antropologii, w której instrumentalnie pojęta wiedza staje się elementem produkcji, handlu, środkiem oddziaływania i transformacji środowiska przyrodniczego i społecznego człowieka, zmieniając i zamieniając go samego na istoty i systemy o zupełnie odmiennym charakterze. Istoty, które otoczone i działające w rzeczywistości hybrydalnej posiadać będą swoje hasła, piny, kody przejścia, warunki rejestracji i cyfrowe reprezentacje w świecie nieśmiertelnych awatarów, ale wielce wątpliwe, czy będą to jeszcze byty osobowe¹⁵.
- c) W podobnym kierunku redukowana jest fundamentalna wartość prawdy. Staje się ona jedynie możliwą do uzgodnienia i potwierdzoną posiadaną wiedzą pragmatyczną wizją rzeczywistości, która pozwala

¹⁴ T. Kobierzycki, „*Gnothi seauton*”. *Dusza, charakter, jaźń, osobowość, osoba. Dzieje pięciu pojęć*, Akademia Muzyczna im. F. Chopina w Warszawie, Warszawa 2001.

¹⁵ M. Krąpiec, *Człowiek jako osoba*, Polskie Towarzystwo Tomasza z Akwinu, Lublin 2005.

na skuteczne osiągnięcie utylitarnych celów. Jeśli zadań tych nie wypełnia, zostaje zweryfikowana negatywnie, obalona, a następnie odrzucona jako rodzaj epistemologicznego, następnie metodologicznego, a w końcu prakseologicznego błędu. Prawda jest w tym ujęciu jedynie projekcją inteligencji na konkretnym etapie jej rozwoju i powinna być przedmiotem modelowania zgodnie z jej aktualnym potencjałem poznawczych możliwości człowieka. Funkcję kryterium prawdziwości spełnia zastosowana metodologia. Rozpoznanie jej uwikłania w sterujące ją wartości poddaje się analogicznej redukcji, usuwając na ogół z pola krytycznej świadomości¹⁶.

Redukcja wartości autotelicznych, jakie przynosi epoka sztucznej inteligencji, niesie ze sobą nie tylko konsekwencje teoretyczne. Równie ważne są konsekwencje praktyczne, głęboko wnikające w realia życia codziennego. Są to:

1. Nasilająca się nierównowaga odmiennych typów wartości. Przejawia się ona w gloryfikacji, a wręcz mitologizacji wartości utylitarnych pozwalających na optymalizację działań wspomaganą sztuczną inteligencją. Powoduje to swoisty imperializm strategii obliczonych na maksymalizację kryteriów skuteczności i efektywności mierzonych za pomocą jedynie parametrycznych wskaźników sukcesu. Tendencje te wyraźnie widać w ewaluacji działań edukacyjnych, zaniedbujących czy wręcz eliminujących kryteria wychowawcze i rozwojowe. Znajduje to swoje patologiczne wręcz skutki w różnego rodzaju „punktozie” i technokratyzacji polityki naukowej, w zaniedbywaniu kulturotwórczych funkcji środowisk szkolnych i akademickich, traktowanych jako korporacyjni producenci wiedzy, w „menedżeryzacji” polityki społecznej i kadrowej organizacji pozarządowych, w formalizacji sposobów podejmowania decyzji, nawet na najniższych poziomach zarządzania. Nacisk na przydatność, a nawet służebność środowisk naukowych w stosunku do swego otoczenia społecznego redukuje fundamentalne różnice odmiennego typu i odmiennej funkcji wiedzy oraz sposobu jej użycia. Kuriozalnym błędem zacierania tego typu różnic jest sprowadzanie nauk przyrodniczych, technicznych, społecznych czy humanistycznych do wspólnego, utylitarnego

¹⁶ H. Mamzer, *Poczucie bezpieczeństwa ontologicznego. Uwarunkowania społeczno-kulturowe*, Wydawnictwo Naukowe UAM, Poznań 2008.

mianownika. Unifikacja aksjologiczna, jaką się narzuca w tak promowanych praktykach, znajduje swe odbicie w tworzonych w tym celu instytucjach, procedurach i podporządkowanych im sposobach finansowania i wspierania nauki. Na tej redukcji oparta jest koncepcja tzw. parametryzacji działania i efektów pracy uczelni, dorobku ich pracowników, a w końcu polityki ich awansowania.

2. Niekontrolowany zalew manierycznie tworzonych metafor, jak: „humanistyka cyfrowa”, „postczłowiek”, „pedagogika sztucznej inteligencji”, sugerujących podmiotowy charakter artefaktów technologicznych. W istocie wszystkie te terminy dokonują metaforycznej redukcji, a następnie wspomnianej już zamiany narzędzi i środków na przedmiot poznania. Nie jest to po prostu humanistyka wspierana cyfrową technologią obróbki i przetwarzania danych, ale już humanistyka cyfrowa. Ostatni z przykładów dokonuje z kolei odwrócenia semantycznego, nadając cechy podmiotowe przedmiotowo i instrumentalnie zorientowanej inteligencji. Gdyby nie sugestywna i władcza siła języka, problem pozostawałby w sferze umowności i operacyjnie dokonywanego skrótu. Niestety jest on nośnikiem daleko idących manipulacji o akceptowanym zawłaszczeniu, skutkującym faktyczną dehumanizacją humanistyki o poważnych konsekwencjach mentalnych, metodologicznych i aksjologicznych. Tysiąc razy powtarzane skróty stają się w ten sposób elementem masowo reprodukowanego kodu tej niby-reprezentacji, iluzorycznego wizerunku prawdy, która staje się zakładnikiem poznawczej i kulturowej mistyfikacji.

3. Algorytmiczny i sterowany charakter działań społecznych. Wartości pozaautoteliczne stają się zasadniczym kryterium zarówno planowania, realizacji, jak i na koniec ewaluacji działań społecznych, realizując wspomniany już kult optymalizacji w każdej niemal sferze życia. Zasady przewidywalności, parametryzacji i kalkulacji stają się przewodnią regułą myślenia o racjonalności, ale i celowości gospodarowania, zarządzania i wartościowania zarówno w sferze działań komercyjnych, jak i publicznych. Filozofię społeczną zastępuje koncepcja, ale i pragmatyka sterowania programem, jaka zdominować ma dialogiczne formy ładu społecznego. Konsekwencją tak obranej strategii jest utylitarna dominanta projektowania treści programów kształcenia, w tym kształcenia uniwersyteckiego, oraz osiągania ich efektów, hedonistyczna wizja jakości życia. Sukces życiowy podporządkowany

jest sukcesowi zawodowemu, a etos wspólnoty podporządkowany potrzebom zależności korporacyjnych. Sztuczna inteligencja stanowi coraz częściej stosowane narzędzie wypracowania wielofunkcyjnych algorytmów, które wyznaczać będą kierunki decyzji organizacyjnych.

Historia polityczna i kulturowa wskazuje jednak, iż największe nawet podporządkowanie wymogom funkcjonowania rynku i struktur politycznych nie eliminuje wartości ukrytych pod powierzchnią adaptacji i konformizmu. W chwilach konfliktu i kryzysu algorytmy sztucznej inteligencji nie są w stanie zapobiec ich „wynurzeniu” oraz ruchom tektonicznym „rewolucji godności” czy też rewitalizacji tożsamości etnicznych lub religijnych po okresie ich zaniku i – wydawałoby się – zapomnienia. To jest właśnie siła i nieredukowalny byt wartości autotelicznych, dopóki człowiek choć w znikomej mierze pozostawi sobie cień swego człowieczeństwa w formie, jaka go obecnie definiuje.

Zakończenie: A jednak pułapka

Czy marzenia pana Twardowskiego miały szansę na realizację? Być może tak, gdyby w celu ich spełnienia użył on sztucznej inteligencji, która ostrzegłaby go przed wariantami błędów popełnionych przez jego inteligencję naturalną. Oznaczałoby to jednak, że tej bardziej efektywnej, nadzorującej formie inteligencji musiałyby być poddane wszelkie tajemnice jego woli, rzeczywiste oczekiwania i kalkulacje, które Twardowski z pewnością chciałby zachować tylko dla siebie. W istocie nie byłaby to już forma inteligencji, ale posiadająca wszelkie ludzkie władze, ambicje i emocje ponadnaturalna forma superosobowości. Wątpliwe, aby Twardowski chciał wydziedziczyć sam siebie i swoją tożsamość aż do tego stopnia, redukując swoje istnienie do innego, podobnego, ale obcego mu bytu. Nie rezygnując jednak z możliwości dokonania wymarzonej transgresji, ufając własnej przebiegłości, popadł w pułapkę. W efekcie nie odzyska już sam siebie, a jego życie, pełne osamotnienia i niemożności, na zawsze będzie już tylko niby-życiem. Z pewnością nie tak je sobie wyobrażał. Przyczyna tej porażki tkwi jednak nie tyle w brakach jego inteligencji i wykalkulowania, ile w arsenale pożądanych wartości, charakterze i rodzaju środków, którymi chciał się posługiwać,

oddziałując na życie swoje i innych, na ukształtowane z nimi relacje i model praktykowanej przez nich duchowości.

Sztuczna inteligencja i stwarzana przez nią obietnica spełniają wszystkie warunki umowy Twardowskiego z Mefistofeilesem. Już dzisiaj wiemy, jakie wyznacza warunki porządku społecznego, jak pozycjonuje człowieka i jego władzę nad realnym sposobem jego życia. Konieczna jest krytyczna, ale i przekładana na konkretne działania refleksja o wartościach, które w tym nowym sposobie życia zostaną uznane za ważne i za dysfunkcjonalne; o tym, jak będą różnicować od dawna definiowany i pożądlivie wymuszany postęp.

Bibliografia

- Gudykunst W. B., Shapiro R. P., *Communication in everyday interpersonal and intergroup encounters*, „International Journal of Intercultural Research” 1996, 20/1, s. 19–45.
- Jaskuła S., *Bezpieczeństwo kulturowe w świecie hybrydalnym. Aspekty teoretyczne i metodologiczne*, w: *Granice tożsamości nauk o bezpieczeństwie. Perspektywa materialna i formalna*, red. T. Kośmider, W. Kitler, Difin, Warszawa 2017, s. 202–218.
- Kłoskowska A., *Koncepcja autoteliczności kultury*, w: *O społeczeństwie i teorii społecznej. Księga poświęcona pamięci Stanisława Ossowskiego*, red. M. Ofierska, Z. Mokrzycki, J. Szacki, PWN, Warszawa 1985.
- Kłoskowska A., *Socjologia kultury*, PWN, Warszawa 1981.
- Kobierzycki T., *„Gnothi seauton”. Dusza, charakter, jaźń, osobowość, osoba. Dzieje pięciu pojęć*, Akademia Muzyczna im. F. Chopina w Warszawie, Warszawa 2001.
- Korporowicz L., *Logotwórcze dynamizmy kultury*, w: tegoż, *Socjologia kulturowa. Kontynuacje i poszukiwania*, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków 2011, s. 15–35.
- Korporowicz L., *Polskie korzenie prawa narodów w budowaniu podmiotowości wspólnot kulturowych*, w: *Eseje o niepodległej*, red. R. Wiśniewski, Narodowe Centrum Kultury, Warszawa 2018, s. 102–129.
- Korporowicz L., *W poszukiwaniu sztucznego człowieczeństwa*, w: *Rzeczywistość hybrydalna. Perspektywa pedagogiczna*, red. S. Jaskuła, Księgarnia Akademicka, Kraków 2023, s. 79–95.
- Krapiec M., *Człowiek jako osoba*, Polskie Towarzystwo Tomasza z Akwinu, Lublin 2005.
- Liotard J. F., *Kondycja ponowoczesna*, tłum. M. Kowalska, J. Migasiński, Aletheia, Warszawa 1997.
- Mamzer H., *Poczucie bezpieczeństwa ontologicznego. Uwarunkowania społeczno-kulturowe*, Wydawnictwo Naukowe UAM, Poznań 2008.
- Orbik Z., *W poszukiwaniu filozoficznych źródeł problematyki sztucznej inteligencji*, „Ethos” 2023, 36, nr 3(143), s. 45–66.

Ossowski S., *U podstaw estetyki*, PWN, Warszawa 1966.

Pink D., *A whole new mind: Moving from the Information Age to the Conceptual Age*, Riverhead Hardcover, New York 2005.

Wiśniewski R., *Transgresja kompetencji międzykulturowych. Studium socjologiczne młodzieży akademickiej*, Wydawnictwo UKSW, Warszawa 2016.

Wójcicki K. W., *Klechdy. Starożytne podania i powieści ludu polskiego i Rusi*, wybór i opracowanie: R. Wojciechowski, Dragon, Bielsko-Biała 2021

Życiński J., *Personalizm po śmierci podmiotu ludzkiego*, w: *Personalizm polski*, red. M. Ru-secki, Wydawnictwo KUL, Lublin 2008.

Artificial intelligence towards autotelic values. Mr. Twardowski today

Summary

The article uses the well-known motif of Mr. Twardowski, who, trusting in his cunning, planned a pact with the devil, thanks to which he was to acquire many goods and skills while controlling the conditions of the pact. In reality, however, he fell into a trap and had to flee alone to the moon. This thread serves to build a thesis on the analogy of the aforementioned pact with the naïve faith of improving man and his abilities through the technological and applied progress of artificial intelligence. The text is an attempt to show man's constant desire for transgression, which brings both positive and negative consequences and unfortunately ends with the loss of his subjectivity exchanged for voluntary alienation and dependence on external factors controlling his life and self-understanding. One of the causes of this failure is the reduction of the human personality, which has mythologized only a fragment of its qualities, which is intelligence, forgetting the others that are not instrumental and that distinguish intelligence from wisdom, autonomy from responsibility, progress from development. The article indicates the theoretical and practical consequences of this type of reductionism and the related dangers that lie in the search for post-man and post-humanism as an ideology of human objectification

Keywords: artificial intelligence, autotelic values, axiological reductionism, objectification technology, transgression

Dylematy etyczne w zastosowaniach sztucznej inteligencji

Niezwykle istotne, a w opinii wielu badaczy zupełnie przełomowe zmiany w funkcjonowaniu współczesnego społeczeństwa niesie ze sobą rozwój i wzrastające spektrum zastosowań sztucznej inteligencji. Choć o intensyfikacji i popularyzacji dyskusji nad jej wpływem na różne sfery życia społecznego można mówić od stosunkowo niedawna, to sam rozwój technologii opartych na algorytmach sztucznej inteligencji zachodzi już od długiego czasu, a rozwiązania tego typu dostępne są wzrastającej liczbie podmiotów. Wiąże się z tym wiele szans, ale także zagrożeń mogących mieć realny wpływ na jakość życia przyszłych pokoleń. W tej sytuacji nie da się uniknąć pytania o etyczny wymiar sztucznej inteligencji. W niniejszym artykule przedstawione zostaną dylematy natury etycznej związane z rozwojem nowych technologii, ze szczególnym uwzględnieniem pojęcia odpowiedzialności człowieka za przyszły kształt społeczeństwa, która w kontekście żywiołowego rozwoju nowych technologii w czasach współczesnych nabiera szczególnej wagi.

¹ Dr hab. Agnieszka Zduniak, prof. UKSW w Warszawie, Instytut Nauk Socjologicznych.

Sztuczna inteligencja – nadzieje i wątpliwości

Żyjemy w epoce, w której sztuczna inteligencja oddziałuje na nasze życie w o wiele większym stopniu, niż jesteśmy w stanie to sobie uzmysłwić. Korzystanie z internetu za pośrednictwem urządzeń osobistych: komputerów, tabletów i telefonów komórkowych, posługiwanie się wyszukiwarkami, obecność w mediach społecznościowych – wszystko to nie tylko modyfikuje sposób komunikowania się ludzi, lecz także wprowadza do ich życia zupełnie nowy element: maszynę, która z narzędzia, jakim była dotychczas, zamienia się stopniowo w nowego, względnie autonomicznego partnera interakcji.

Sztuczna inteligencja w coraz większym stopniu wpływa na postrzeganie świata przez człowieka. Dokonuje się to przez odpowiednią selekcję informacji, promowanie określonych treści kosztem innych, kreowanie potrzeb, zainteresowań i poglądów. Działając dyskretnie, w tle, AI wpływa w realny sposób na zachowania i działania ludzi oraz na konkretne decyzje podejmowane przez nich w sferze polityki, ekonomii, a także w sferze prywatnej. Nie ma wątpliwości w kwestii tego, że jej wpływ na życie ludzi na całym świecie będzie wzrastał, co z czasem doprowadzi do trwałych zmian w życiu społecznym.

Entuzjaści nowych technologii dostrzegają przede wszystkim płynące z jej zastosowania korzyści, których rzeczywistość jest bardzo dużo. Możliwy do wykorzystania potencjał AI w zakresie poprawy warunków ludzkiego życia wydaje się wręcz nieograniczony. Specjaliści w poszczególnych dziedzinach wynajdują coraz to nowe jej zastosowania w ratowaniu ludzkiego zdrowia, usprawnianiu procesu produkcji przez automatyzację wielu czynności, ochronie przyrody, przetwarzaniu ogromnych ilości danych, ułatwieniach w komunikacji i wielu innych. Koncentracja na niewątpliwych korzyściach, jakie rozwój AI może przynieść ludzkości, przesłania jednak często pytania natury etycznej o to, jakie długofalowe skutki może mieć rozwój nowych technologii i czy rzeczywistość na dłuższą metę będą one służyć dobru człowieka.

Przyjmowane często bezrefleksyjnie założenie o etycznej neutralności techniki nie da się obronić, jeśli przyjrzeć się potencjalnym negatywnym skutkom niepowstrzymanego rozwoju nowych technologii

bez oglądania się na ich konsekwencje. Podobnie jak i inne osiągnięcia nauki – te dobrze znane i te najnowsze – sztuczna inteligencja może być wykorzystana niekoniecznie do tych celów, do jakich została stworzona. Jeśli przyjrzeć się choćby sferze komunikowania, najbardziej chyba kojarzącej się z zastosowaniem algorytmów AI, to bez trudu można zauważyć, że mogą one służyć informowaniu, ale i celowemu szerzeniu dezinformacji; nawiązywaniu i podtrzymywaniu kontaktów, ale także wykluczaniu osób i całych grup społecznych (hejt internetowy, mowa nienawiści, *cancelling*). Mogą powiększać sferę ludzkiej wolności, a jednocześnie umożliwiać kontrolę i inwigilację na niespotykaną dotąd skalę; są w stanie odpowiadać na autentyczne ludzkie potrzeby, lecz również świadomie kreować potrzeby nowe i często szkodliwe, ale za to korzystne np. dla wielkich koncernów, dla których zysk, opłacalność i potencjał komercyjny stanowią nierzadko cele same w sobie. To samo można powiedzieć o innych obszarach zastosowania sztucznej inteligencji: w sferze gospodarki, edukacji, medycyny, polityki i wielu jeszcze innych obszarach zastosowań może ona służyć ludziom, ale także zwrócić się przeciwko człowiekowi – szczególnie w sytuacjach, gdy jej użytkownicy zaczną przedkładać doraźne cele ekonomiczne czy polityczne nad szeroko rozumiane dobro wspólne. Postęp technologiczny i etyka są zatem ze sobą ściśle powiązane².

Dlatego też niezwykle istotną kwestią jest rozumienie etycznych dylematów rozwoju i zastosowania sztucznej inteligencji, świadomość związanych z nimi szans i zagrożeń. Może się do tego przyczynić inicjowanie i podtrzymywanie dyskusji nad rozumieniem moralności w dobie rozwoju nowych technologii, w których winni brać udział nie tylko eksperci, jak na ogół odbywało się to dotychczas, ale także szerokie kręgi społeczne. To właśnie od refleksji i działań na rzecz świadomego i etycznego kształtowania technologii opartej na sztucznej inteligencji zależeć będzie kształt życia społecznego w najbliższej i dalszej przyszłości.

² P. G. Kirchschräger, *The Correlation between Ethics and Technology*, w: *Religion in the Age of Digitalization: From New Media to Spiritual Machines*, ed. G. Isetti, E. Innerhofer, H. Pechlaner, M. de Rachewiltz, Routledge, Abingdon 2020.

Czy sztuczna inteligencja może być moralna?

Nowe technologie oparte na sztucznej inteligencji pod wieloma względami przewyższają możliwości ludzkiego umysłu. W ułamku sekundy są w stanie przetwarzać niewiarygodne ilości danych, tworzyć teksty, korygować błędy, opracowywać rozwiązania skomplikowanych problemów, które bez ich zastosowania wymagałyby wytężonej i długotrwałej pracy dużych zespołów ludzkich. Mimo godnych podziwu możliwości kognitywnych AI nie ma właściwości ludzkiej inteligencji, choć w wielu aspektach jest w stanie ją imitować. Tym, co stanowi integralną część ludzkiego umysłu, są także emocje, rozumienie kontekstu społecznego podejmowanych działań, inteligencja emocjonalna, myślenie twórcze. Tylko człowiek dysponuje intuicją moralną i jest w stanie rozstrzygać etyczne dylematy, co nie zawsze dokonuje się w oparciu o logiczne rozumowanie. Wielu przedstawicieli filozofii, etyki i psychologii opowiada się raczej za teorią, według której osąd moralny dokonuje się bardziej na podstawie przeżyć – emocji, podpowiadających człowiekowi, że jakieś działanie jest moralnie słuszne lub przynajmniej słuszniejsze niż działanie alternatywne, choćby rozumowanie bazujące na zasadach logiki temu przeczyło. Dylemat etyczny to „konieczność dokonania trudnego wyboru, podjęcia kontrowersyjnej decyzji w sytuacji niejednoznacznej. Zazwyczaj też dylematy pojawiają się tam, gdzie spontaniczny, «ludzki» ogląd jakiejś sprawy styka się z wymiarem instytucjonalnym i mechanicznie działającą normą”³. To właśnie algorytm, owa „mechanicznie działająca norma”, uniemożliwia systemom bazującym na AI podejmowanie decyzji, które w ludzkim rozumieniu mają charakter decyzji moralnych.

Przyjrzyjmy się jeszcze dwóm elementom istotnym w dyskusji o etycznym wymiarze jakiegokolwiek działania: pojęciom wolności oraz odpowiedzialności. Decyzje i podjęte w ich wyniku czyny lub dopuszczone zaniechania mogą podlegać ocenie moralnej tylko wówczas, gdyby zostały podjęte w sposób wolny. Co oczywiste, wolność czy też autonomia nie są pojęciami, które można by łączyć z najbardziej nawet zaawansowanymi systemami technologicznymi, chociaż terminów tych używa się niekiedy

³ D. Bąk, *Pojęcie dylematu etycznego. Zarys problemu dylematu etycznego w organizacji*, „Prakseologia” 2013, nr 154, s. 5–6.

w języku marketingu (np. termin „samochód autonomiczny”). Umiejętność samodzielne, autonomiczne podejmowania decyzji to cechy immanentnie ludzkie. Hierarchie wartości, uznawane normy, obyczaj i moralność, które zawsze wpływają w jakimś stopniu na ludzkie działania, ukierunkowują je i legitymizują, nie mają jednak znaczenia w funkcjonowaniu sztucznej inteligencji, ponieważ jej działaniem kieruje algorytm. On z kolei jest zależny od „pierwszej linijki kodu, który zawsze pochodzi od człowieka”⁴. O działaniach sztucznej inteligencji, także tych, które mają jakiś wymiar moralny, zawsze decyduje zatem człowiek, który zaprogramował jej funkcjonowanie; na nim także spoczywa odpowiedzialność za skutki owych działań. Pojęcie odpowiedzialności staje się więc kluczowym elementem rozważań o etycznym wymiarze wdrażania nowych technologii.

Odpowiedzialność jako kategoria etyczna

Koncepcja odpowiedzialności stanowi istotny aspekt wielu rozważań filozoficznych i etycznych. Roman Ingarden wskazywał, że o fenomenie tym można mówić wówczas, gdy ktoś ponosi za coś odpowiedzialność, podejmuje za coś odpowiedzialność, jest pociągany do odpowiedzialności, a także gdy działa w sposób odpowiedzialny. Jednak o odpowiedzialności mówić można tylko w przypadkach, gdy człowiek – osoba podejmuje decyzje i działa w sposób wolny. Ograniczenie wolności lub jej brak wyklucza zatem odpowiedzialność. Istotne jest tu rozróżnienie dwóch aspektów: pierwszy z nich – bardziej oczywisty – to odpowiedzialność „po fakcie”, a więc po podjęciu decyzji i dokonaniu danego czynu; drugi, nie mniej istotny, to odpowiedzialność poprzedzająca czyn i jego skutki, antycypująca. Odpowiedzialność tego typu obejmuje przewidywanie konsekwencji czynów jeszcze niedokonanych i decyzji, które jeszcze nie zapadły⁵. Człowiek może także,

⁴ P. G. Kirchschräger, *Künstliche Intelligenz und Ethik*, „Kirche und Gesellschaft” 2021, Nr. 478, s. 3–16.

⁵ R. Ingarden, *Książeczka o człowieku*, Wydawnictwo Literackie, Kraków 2003, s. 77, 133. Tenże, *Wykłady krakowskie. Wykład trzynasty*, w: R. Ingarden, *Wykłady z etyki*, PWN, Warszawa 1989, s. 296.

korzystając ze swojej wolności, odmówić wzięcia odpowiedzialności za swoje działania, nie oznacza to jednak, że owa odpowiedzialność przestaje istnieć.

Na wzajemne powiązanie wolności i odpowiedzialności zwracali uwagę Emmanuel Lévinas oraz Hans Jonas. Dla Lévinasa odpowiedzialność stanowi fundament ludzkiej podmiotowości; w jego ujęciu oznacza ona radykalne poświęcenie się drugiemu człowiekowi⁶. Zarówno odpowiedzialność, jak i wolność są zawsze powiązane z wartościami i powinny służyć ich urzeczywistnianiu. W koncepcji Jonasa zawarty został postulat przejścia od etyki antropocentrycznej do etyki biocentrycznej; ludzka odpowiedzialność winna obejmować nie tylko innych ludzi czy nawet całą ludzkość, lecz także przyrodę, która jest niezbędnym warunkiem istnienia i trwania gatunku ludzkiego⁷. W swoim najstynniejszym dziele *Zasada odpowiedzialności. Etyka dla cywilizacji technologicznej* niemiecki filozof przedstawia filozoficzne podstawy odpowiedzialności za losy świata (w tym wszystkich bytów ożywionych), przeciwstawiając się koncepcji neutralności aksjologicznej nauki i techniki. Umożliwiają one postęp, ale ich nieodpowiedzialne wykorzystywanie może zagrażać samemu istnieniu człowieka.

Karol Wojtyła w swoim studium *Osoba i czyn* stwierdza również, że istotą odpowiedzialności jest wolność: „Człowiek dlatego bywa odpowiedzialny za swe czyny i dlatego przeżywa odpowiedzialność, że ma zdolność odpowiadania wołaniu wartości”⁸. Odpowiedzialność może się odnosić do podmiotu – mówimy wówczas o odpowiedzialności za siebie i przed sobą samym – ale może być także odpowiedzialnością przed innymi ludźmi, co wynika z przyporządkowania człowieka jako osoby do całego świata osób, mającego swoją strukturę międzyludzką oraz strukturę społeczną. W ramach struktury religijnej odpowiedzialność „przed” oznacza także odpowiedzialność przed Bogiem. Zdaniem Wojtyły jest ona więc w nierozłączny sposób powiązana z osobą jako podmiotem i jednocześnie z ludzkim sumieniem⁹. Również Józef Tischner podkreśla, iż doświadczenie odpowiedzialności u człowieka

⁶ E. Lévinas, *Całość i nieskończoność. Esej o zewnętrzności*, tłum. M. Kowalska, PWN, Warszawa 1998, s. 208–209.

⁷ H. Jonas, *Zasada odpowiedzialności. Etyka dla cywilizacji technologicznej*, tłum. M. Klimowicz, Platan, Kraków 1996, s. 168.

⁸ K. Wojtyła, *Osoba i czyn oraz inne studia antropologiczne*, WN KUL, Lublin 2000, s. 213.

⁹ Tamże, s. 215.

możliwe jest dlatego, że jest on istotą wolną i jednocześnie zdolną do wybierania spośród wielu wartości, tych, które mają charakter absolutny i nierelatywny¹⁰.

Przytoczone przykłady rozumienia i stosowania koncepcji odpowiedzialności nie pozostawiają wątpliwości co do tego, że nawet najsprawniej działająca sztuczna inteligencja nie stwarza warunków do zaistnienia w niej samej fenomenu odpowiedzialności. Nie ma ona bowiem kluczowych do tego cech, którymi są: wolność, osobowa tożsamość, sumienie, umiejętność odróżnienia wartości absolutnych od relatywnych, zdolność do tworzenia relacji z ludźmi, przyrodą i Bogiem. Nauka i technika, pozbawione ograniczeń wynikających z odpowiedzialności za losy ludzi, mogą łatwo wymknąć się spod kontroli, a skutki tego mogą być trudne do przewidzenia. Odpowiedzialność za ich rozwój spoczywa tylko i wyłącznie na człowieku, który wciąż na nowo musi zadawać sobie pytanie o to, czy czyni wszystko, aby jego wynalazek nie okazał się dla ludzkości puszką Pandory.

Współczesna dyskusja na temat sztucznej inteligencji

Brytyjski astrofizyk Stephen Hawking nie krył się ze swoimi obawami, że sztuczna inteligencja może oznaczać koniec ludzkości.

Jeśli nie przygotujemy się na potencjalne ryzyko i nie nauczymy się go unikać, sztuczna inteligencja może stać się najgorszym wydarzeniem w historii naszej cywilizacji. Niesie ona ze sobą zagrożenia, takie jak potężne, autonomiczne rodzaje broni czy nowe sposoby, dostępne nielicznym, aby podporządkować sobie całą resztę¹¹.

Na zagrożenia związane z rozwojem nowych technologii zwraca uwagę również Elon Musk: „Sztuczna inteligencja jest o wiele bardziej niebezpieczna niż broń atomowa. Dlaczego zatem nie ma nad nią nadzoru?”¹².

¹⁰ J. Tischner, *Myslenie według wartości*, Znak, Kraków 1993, s. 49.

¹¹ S. Hawking, cyt. przez A. Kharpal, *Stephen Hawking says A.I. could be „worst event in the history of our civilization”*, CNBC News, 6.11.2017.

¹² E. Musk, cyt. przez C. Clifford, *Elon Musk: „Mark my words – A.I. is far more dangerous than nukes”*, CNBC News, 13.03.2018.

Sami twórcy sztucznej inteligencji przyznają, że jest ona najważniejszą z rozwijanych obecnie na świecie technologii i że prawdopodobnie „zmieni nasze podejście do technologii. Zmieni nasze podejście do drugiego człowieka [...] zmieni nawet nasze postrzeganie istoty człowieczeństwa”¹³.

Jednym z zagrożeń, jakie niesie dla człowieka wdrażanie systemów opartych na sztucznej inteligencji, jest pokusa rozproszenia odpowiedzialności, „dzielenia” jej ze sztuczną inteligencją czy nawet pozbywania się odpowiedzialności za działania nieetyczne podjęte przez „bezduszny” system. Twórcy sztucznej inteligencji zdają sobie sprawę z tego, że może dojść do sytuacji, w których „autonomiczny” robot zacznie podejmować decyzje sprzeczne z celami, do których został zaprogramowany. Pojawiła się w związku z tym kontrowersyjna propozycja nadania sztucznej inteligencji osobowości prawnej. Koncepcja ta byłaby korzystna dla producentów systemów opartych na AI, którzy w taki sposób unikaliby odpowiedzialności za ich działania, jednak z punktu widzenia prawa, a w szczególności etyki, jest nie do przyjęcia. O tym, że systemy oparte na sztucznej inteligencji mogą wymknąć się spod kontroli, przekonali się już twórcy Facebooka, którzy w ramach udoskonalania jego funkcji stworzyli dwa boty – Alice i Boba – porozumiewające się z użytkownikami i między sobą w języku angielskim. Po pewnym czasie zostały wyłączone, ponieważ zaczęły komunikować się między sobą w nowym języku, niezrozumiałym dla nikogo oprócz nich samych¹⁴. Stephen Hawking, mówiąc o zagrożeniach ze strony sztucznej inteligencji, zwracał uwagę właśnie na to, że gdy stanie się ona ogólnodostępna – czyli przestanie się uczyć w zamkniętym środowisku – może całkowicie wymknąć się spod kontroli człowieka.

Raport stworzony w 2020 r. przez amerykański ośrodek Pew Research Center oraz Imagining the Internet Center, jednostkę naukową przy Uniwersytecie w Elon w Północnej Karolinie, również nie napawa optymizmem. W badaniu wzięły udział 602 osoby będące ekspertami z różnych dziedzin

¹³ S. Sheard, *Sztuczna inteligencja to najważniejsza z obecnie rozwijanych na świecie technologii*, 2016, <https://businessinsider.com.pl/technologie/nowe-technologie/sztuczna-inteligencja-to-najwazniejsza-rozwijana-obecnie-technologie/qscny5j> (28.01.2024).

¹⁴ P. Stylec-Szromek, *Sztuczna inteligencja – prawo, odpowiedzialność, etyka*, „Zeszyty Naukowe Politechniki Śląskiej” 2018, z. 123, s. 504.

związanych z rozwojem technologii – prawnicy, liderzy biznesu, twórcy nowych rozwiązań technologicznych, informatycy, naukowcy i aktywiści. Najważniejsze pytanie w badaniu brzmiało następująco: „Czy w roku 2030 większość rozwiązań AI, stosowanych we wszelkiego rodzaju organizacjach, będzie funkcjonować zgodnie z etycznymi zasadami służącymi przede wszystkim dobru wspólnemu?”. Większość badanych (68%) wybrała odpowiedź *takie standardy nie będą wtedy obowiązywać*. Przeciwnego zdania było jedynie 32% respondentów. Wskazywali oni przede wszystkim na to, że sztuczna inteligencja może bez ograniczeń reprodukcować kulturowe wzory dyskryminacji, co będzie się przyczyniać raczej do utrwalania niesprawiedliwości niż do jej zniesienia. Należałoby zatem od początku powstawania każdego systemu implementować w nim zasady etyczne, jednak nie jest to łatwe w sytuacji braku powszechnego konsensusu co do rozumienia dobra i zła¹⁵.

W jaki sposób zatem zaradzić takiej sytuacji, mając świadomość, że rozwoju nowych technologii nie da się już powstrzymać? Odpowiedzią mogą być realizowane w różnych ośrodkach naukowych programy badawcze, których twórcy, uznając fakt, że nie można pozwolić sztucznej inteligencji decydować o losach ludzi, ponieważ nie jest ona zdolna ani do samodzielnego rozstrzygnięcia dylematów moralnych, ani do odczuwania wyrzutów sumienia, mają jednak nadzieję, że będzie można zaprogramować ją w taki sposób, aby nie podejmowała decyzji szkodliwych dla ludzi.

Przykładem takiego programu jest projekt *Moral Choice Machine* realizowany przez uniwersytet w Darmstadt¹⁶. Jego twórcy są zdania, że sztuczna inteligencja może uczyć się moralności nie w sensie autonomicznych decyzji moralnych, lecz raczej na zasadzie przyswajania wiedzy o tym, co jest moralne w ludzkim rozumieniu. Punkt wyjścia stanowi fakt, że AI potrafi wyszukiwać w tekście informacje i łączyć je ze sobą. Uczestnicy projektu zdefiniowali najpierw dobro i zło, a następnie wgrali do systemu liczne teksty kultury pochodzące z różnych epok. Zadaniem sztucznej inteligencji było

¹⁵ E. Nosarzewska, *Czy sztuczna inteligencja może być etyczna?*, Raport Pew Research Center 2020, https://ptsp.pl/sztuczna-inteligencja-etyka/?gclid=EAIaIQobChMIrP7arPXIgaMVEV8YCh-3-fAj3EAAYBCAAEgKDLPD_BwE (14.12.2023).

¹⁶ C. Fron, *Künstliche Intelligenz: Moral ohne Bewusstsein*, Deutschlandfunk 10.08.2019, <https://www.deutschlandfunk.de/kuenstliche-intelligenz-moral-ohne-bewusstsein-100.html> (21.12.2023).

przyporządkowanie konkretnych haseł z tekstów do kategorii „dobra” lub „zła”. W takim kluczu sztuczna inteligencja mogła dokonywać nieskomplikowanych wyborów moralnych. Jedna z uczestniczek projektu podsumowała jego efekty następująco: „Można teraz postawić systemowi pytanie, w jakim stopniu pożądane są konkretne czynności. System decyduje, czy ma do czynienia z czynnością, którą należałoby podjąć, czy z taką, z której należy raczej zrezygnować, oceniając je według ogólnych kryteriów dobra i zła (...) Decyzje, które podejmuje maszyna, to nie są wybitne osiągnięcia filozoficzne; to raczej proste reguły, które mówią, że lepiej jest nie zabijać i dobrze jest kochać swoich rodziców”¹⁷.

Naukowcy są jednak sceptyczni co do możliwości skonstruowania kiedykolwiek programu komputerowego będącego w stanie podejmować decyzje moralne. Algorytmy mają potencjał uczenia się funkcjonujących w społeczeństwie norm i wartości, które mogą stanowić dla nich coś w rodzaju kompasu moralnego; ich „decyzje” będą jednak tylko odzwierciedleniem pewnych wyobrażeń na temat tego, jak w konkretnej sytuacji postąpiłby ludzie. Brak ludzkiej świadomości nie pozwala im na podejmowanie prawdziwych decyzji moralnych. Jediną instancją moralną dla sztucznej inteligencji pozostaje zatem człowiek, który ją programuje. Należy przy tym pamiętać, że decyzje algorytmów opierają się na wszystkich danych, którymi są zasilane. Znajdują się wśród nich także informacje historyczne, odzwierciedlające zachowania społeczne w określonym przedziale czasu. Ukryte w nich komponenty społeczne – wartości i opinie, także te niepożądane, zawierające np. stereotypy czy uprzedzenia – stają się częścią sztucznej inteligencji.

Zakończenie

Bez względu na stopień rozwoju sztucznej inteligencji odpowiedzialność za jej działanie spoczywa wyłącznie na ludziach. To oni będą definiować zasady etyczne, ustalać hierarchię wartości, czuwać nad przestrzeganiem norm moralnych i prawnych. Transformacja cyfrowa nie może się odbywać

¹⁷ Tamże.

bez jakichkolwiek ograniczeń. Twórcy i użytkownicy nowych technologii powinni wciąż na nowo zadawać sobie pytanie o to, co właściwie chcemy osiągnąć przez ich stosowanie. Już w tej chwili da się zauważyć wiele kwestii kontrowersyjnych, które prędzej czy później będą wymagać regulacji prawnych. Rozwój technologii cyfrowych daje ogromne możliwości, zwiększa efektywność i pomnaża zyski, ale nie można go utożsamiać z interesem społecznym rozumianym jako działanie na rzecz dobra wspólnego. Dlatego w dyskurs na temat granic postępu technicznego muszą się włączyć, oprócz przedstawicieli nauki i techniki, również specjaliści z zakresu etyki, prawa, filozofii i ekonomii. Niezwykle ważna jest również edukacja całego społeczeństwa na temat szans i zagrożeń związanych z technologiami cyfrowymi. Konieczne zmiany, ograniczenia i regulacje prawne mogą się dokonać jedynie pod naciskiem opinii publicznej i świadomych użytkowników.

Bibliografia

- Bąk D., *Pojęcie dylematu etycznego. Zarys problemu dylematu etycznego w organizacji*, „Prakseologia” 2013, nr 154, s. 5–32.
- Clifford C., *Elon Musk: „Mark my words – A.I. is far more dangerous than nukes”*, CNBC News, 13.03.2018.
- Fron C. von, *Künstliche Intelligenz: Moral ohne Bewusstsein*, Deutschlandfunk 10.08.2019, <https://www.deutschlandfunk.de/kuenstliche-intelligenz-moral-ohne-bewusstsein-100.html> (21.12.2023).
- Ingarden R., *Książeczka o człowieku*, Wydawnictwo Literackie, Kraków 2003.
- Ingarden R., *Wykłady krakowskie. Wykład trzynasty*, w: R. Ingarden, *Wykłady z etyki*, PWN, Warszawa 1989.
- Jonas H., *Zasada odpowiedzialności. Etyka dla cywilizacji technologicznej*, tłum. M. Klimowicz, Platan, Kraków 1996.
- Kharpal A., *Stephen Hawking says A.I. could be „worst event in the history of our civilization”*, CNBC News, 6.11.2017.
- Kirchschläger P. G., *Künstliche Intelligenz und Ethik*, „Kirche und Gesellschaft” 2021, Bd. 478, s. 3–16.
- Kirchschläger P. G., *The Correlation between Ethics and Technology*, w: G. Isetti, E. Innerhofer, H. Pechlaner, M. de Rachewiltz (ed.), *Religion in the Age of Digitalization: From New Media to Spiritual Machines*, Routledge, Abingdon 2020.
- Lévinas E., *Całość i nieskończoność. Esej o zewnętrżności*, tłum. M. Kowalska, PWN, Warszawa 1998.

- Nosarzewska E., *Czy sztuczna inteligencja może być etyczna?*, Raport Pew Research Center 2020, https://ptsp.pl/sztuczna-inteligencja-etyka/?gclid=EAIAIQobChMIrP7arPXI-gAMVeV8YCh3-fAj3EAAYBCAAEgKDLPD_BwE (14.12.2023).
- Shead S., *Sztuczna inteligencja to najważniejsza z obecnie rozwijanych na świecie technologii*, 2016, <https://businessinsider.com.pl/technologie/nowe-technologie/sztuczna-inteligencja-to-najwazniejsza-rozwijana-obecnie-technologie/qscny5j> (28.01.2024).
- Stylec-Szromek P., *Sztuczna inteligencja – prawo, odpowiedzialność, etyka*, „Zeszyty Naukowe Politechniki Śląskiej” 2018, z. 123, s. 501–509.
- Tischner J., *Myslenie wedlug wartosci*, Wydawnictwo Znak, Kraków 1993.
- Wojtyła K., *Osoba i czyn oraz inne studia antropologiczne*, WN KUL, Lublin 2000.

Ethical dilemmas in the applications of artificial intelligence

Summary

The text is devoted to the opportunities and threats resulting from the development and application of artificial intelligence. The development of new technologies increasingly affects the functioning of humans and society and may also have a huge impact on the quality of life of future generations. In this situation, it is important to reflect on the ethical dimension of artificial intelligence. The main axis of the author's considerations is the concept of responsibility in philosophical reflection and its contemporary application to the changes taking place in society as a result of the development of digital technologies.

Keywords: ethics and artificial intelligence, moral dilemmas, responsibility

Zagrożenia wynikające z nieetycznego i nieuregulowanego prawnie zastosowania technologii sztucznej inteligencji

Wprowadzenie

Sztuczna inteligencja jest dziedziną informatyki, w ramach której budowane są coraz doskonalsze maszyny i komputery z możliwością przeprowadzania operacji logicznych. Obecnie systemy sztucznej inteligencji mogą wykonywać zadania, które do niedawna był w stanie realizować tylko człowiek, wykorzystujący własną inteligencję. Do takich zadań, którym potrafi już podołać sztuczna inteligencja, zalicza się rozpoznawanie mowy, podejmowanie decyzji, percepcję wzrokową i tłumaczenie języków. Tego rodzaju złożone zadania może ona wykonywać dzięki wbudowanym w nią algorytmom. Algorytm jest zestawem instrukcji opisujących to, w jaki sposób określone zadanie powinno być wykonane. Sposób działania algorytmów opiera się w dużym stopniu także na naśladowaniu określonych powtarzalnych działań ludzi, którzy w swoich codziennych, wyuczonych i powtarzalnych czynnościach czy podczas rozwiązywania określonych zadań stosują swego rodzaju

¹ Dr Dariusz Prokopowicz, Wydział Społeczno-Ekonomiczny UKSW.

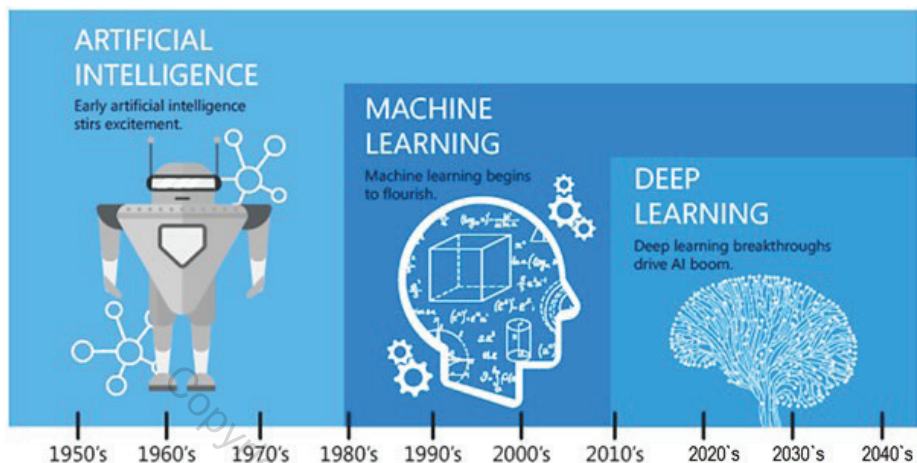
nieformalne algorytmy itd. Przewaga wyposażonych w sztuczną inteligencję komputerów nad człowiekiem wynika z tego, że komputery mogą znacznie szybciej i precyzyjniej przeprowadzać określone czynności zgodnie z ustalonym algorytmem postępowania². Algorytmy wbudowane w systemy operacyjne komputerów, urządzeń mobilnych i innych urządzeń wyposażonych w mikroprocesory są zwykle wykorzystywane do przetwarzania danych, wykonywania mniej lub bardziej złożonych obliczeń i przeprowadzania automatycznego wnioskowania. Uczenie maszynowe należy do najpopularniejszych algorytmów stosowanych w ramach rozwoju zastosowań sztucznej inteligencji³. Proces uczenia się, któremu poddawany jest określony system sztucznej inteligencji, polega na poszerzaniu nowej wiedzy deklaratywnej, rozwijaniu umiejętności poznawczych i/lub motorycznych przez praktykę lub instruktaż. Historia rozwoju tego rodzaju uczenia sięga początków rozwoju komputerów. Jednak dopiero w ostatnich latach, dzięki dokonaniem postępowi technologicznemu, zwiększonej mocy obliczeniowej mikroprocesorów, przetwarzaniu danych w ramach coraz większej pamięci operacyjnej, pracy na coraz większych zbiorach danych zgromadzonych na dyskach o dużej pojemności, rozwój uczenia maszynowego znacznie przyspieszył. Uczenie głębokie natomiast to subsegment problematyki uczenia maszynowego. Uczenie głębokie należy do najszybciej rozwijających się aplikacji sztucznej inteligencji. Przyczyną są duże możliwości zastosowań sztucznej inteligencji wyposażonej w narzędzia zbudowane dzięki uczeniu głębokiemu. Wspomniane możliwości to rozwiązywanie problemów, które do niedawna uważano za niemożliwe do rozstrzygnięcia, zbyt skomplikowane i/lub wiążące się z dużą ilością danych i informacji do przetworzenia. W proces uczenia głębokiego zaangażowane są warstwowe sieci neuronowe w celu rozpoznawania wzorców i złożonych relacji występujących w dużych zbiorach danych. Celem zastosowania uczenia głębokiego należy dysponować ogromnym zestawem danych i potężną mocą obliczeniową mikroprocesorów⁴.

² C. Cath, *Governing Artificial Intelligence: Ethical, legal, and technical opportunities and challenges*, Royal Society, London 2018, s. 37–38.

³ J. Lesandrini, M. Y. Idris, D. S. Reis, *The Ethics of Artificial Intelligence and Machine Learning*, „Journal of Radiology Nursing” 2023, no. 42(3), s. 265–266.

⁴ S. Talari, *Why is Artificial Intelligence Important*, <https://mindmajix.com/why-ai-is-more-important-than-you-think> (17.08.2023).

Rysunek 1. Uczenie głębokie jako poddziedzina uczenia maszynowego w ramach rozwoju zastosowań sztucznej inteligencji



Od wczesnego przypiętywu optymizmu w latach 50. mniejsze podzbiory sztucznej inteligencji – najpierw uczenie maszynowe, a następnie uczenie głębokie, podzbiór uczenia maszynowego – powodowały coraz większe zakłócenia.

Źródło: R. Savaram, *Artificial Intelligence Vs Machine Learning*, <https://mindmajix.com/machine-learning-vs-ai> (17.12.2023)

Powszechnie przyjmuje się, że sztuczna inteligencja jest technologią rozwijającą się w dziedzinie informatyki, polegającą na tworzeniu coraz bardziej zaawansowanych systemów informatycznych, zdolnych inteligentnie wykonywać czynności, które wcześniej przeprowadzali wyłącznie ludzie⁵. Rozwój tej technologii opiera się m.in. na prowadzeniu badań i definiowaniu prawidłowości dotyczących inteligentnych zachowań człowieka, by następnie zastosować je w algorytmach i programach zdolnych uczyć się naśladowania ludzkich zachowań i wykonywania czynności realizowanych do niedawna wyłącznie przez ludzi. Systemy informatyczne wykorzystujące sztuczną inteligencję stosowane są m.in. do: rozpoznawania tekstów, obrazów, tworzenia translatorów, gier symulacyjnych, zaawansowanych symulatorów złożonych procesów i programów rozwiązujących w sposób inteligentny złożone zadania⁶. Określenie **sztuczna**

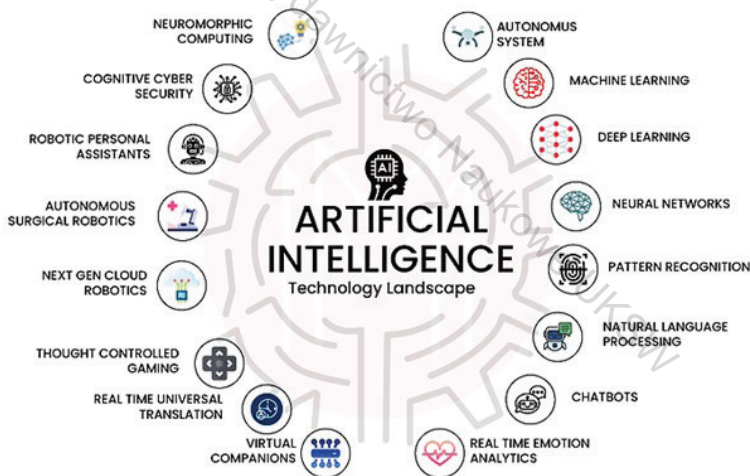
⁵ P. Stylec-Szromek, *Sztuczna inteligencja...*, art. cyt., s. 503.

⁶ K. Różanowski, *Sztuczna inteligencja: rozwój, szanse i zagrożenia*, „Zeszyty Naukowe Warszawskiej Wyższej Szkoły Informatyki” 2007, nr 2, s. 115–116.

inteligencja zostało po raz pierwszy zastosowane przez Johna McCarthy'ego w Dartmouth w 1955 r., podczas konferencji, która dotyczyła tej problematyki⁷.

Skala rozwoju nowych zastosowań technologii sztucznej inteligencji jest tak szeroka, że zmiany generowane przez wdrażanie tej technologii w różnych sektorach gospodarki i różnych sferach życia człowieka są porównywane do wynalezienia elektryczności i kluczowej roli jej zastosowań w okresie drugiej rewolucji technologicznej czy nawet do opanowania przez człowieka pierwotnego zdolności posługiwania się ogniem⁸. Na poniższym rysunku przedstawiono główne rodzaje rozwiązań technologicznych opartych na sztucznej inteligencji. Owe rozwiązania rozwijają się w sposób ściśle skorelowany z dokonującym się w ostatnim czasie szybkim postępem zastosowań technologii AI w podmiotach gospodarczych, instytucjach finansowych i publicznych, instytucjach badawczo-rozwojowych i innych podmiotach funkcjonujących w różnych sektorach gospodarki.

Rysunek 2. Główne rodzaje rozwiązań technologicznych opartych na sztucznej inteligencji



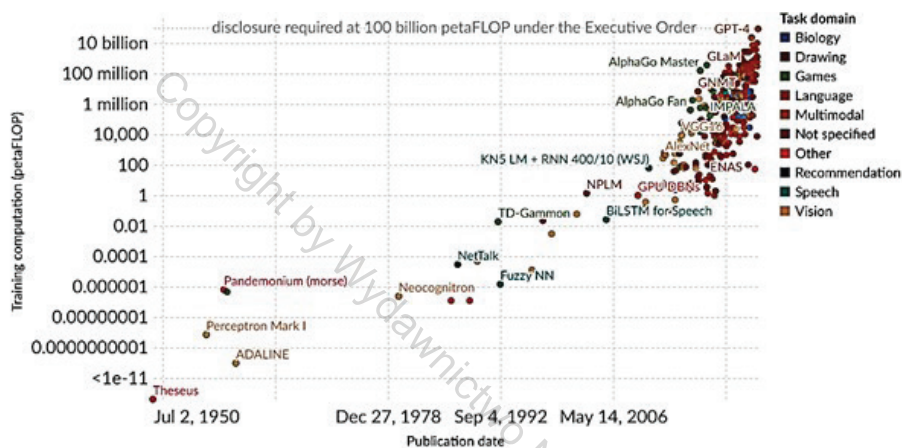
Źródło: M. Yerukala, *Artificial Intelligence Tutorial*, <https://mindmajix.com/artificial-intelligence-tutorial> (16.12.2023)

⁷ J. McCarthy, P. J. Hayes, *Some philosophical problems from the standpoint of artificial intelligence*, w: *Readings in artificial intelligence*, ed. B. L. Webber, N. J. Nilsson, Morgan Kaufmann Publishers 1969, s. 431–450.

⁸ M. Bieroński, *Etyczne i moralne wyzwania związane ze stosowaniem Sztucznej Inteligencji*, „Kieleckie Studia Teologiczne” 2020, nr 19, s. 11–12.

Dynamikę rozwoju technologii sztucznej inteligencji i jej zastosowań przedstawia poniższy wykres, na którym kwestię tego rozwoju odniesiono do wzrostu skali mocy obliczeniowej procesów wykorzystywanych do szkolenia znanych systemów sztucznej inteligencji. Analizę przeprowadzono dla okresu 1950–2023, z uwzględnieniem kluczowych dziedzin zastosowań technologii AI.

Wykres 1. Obliczenia wykorzystywane do szkolenia znanych systemów sztucznej inteligencji (1950–2023)



Obliczenia mierzy się w całkowitym petaFLOP, czyli 10^{15} operacji zmiennoprzecinkowych oszacowanych na podstawie literatury dotyczącej sztucznej inteligencji.

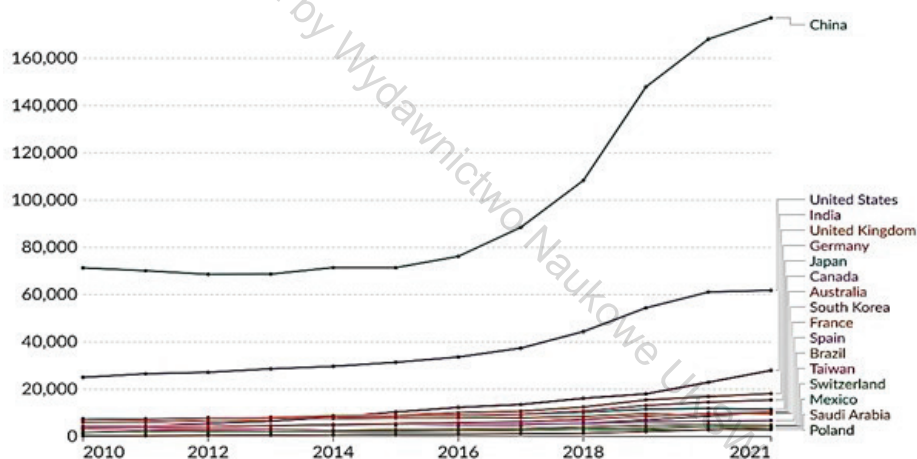
Źródło: Interactive Charts on Artificial Intelligence, <https://OurWorldInData.org/artificial-intelligence> (9.01.2024)

Sztuczna inteligencja, czy tego chcemy czy nie, już funkcjonuje w różnych sferach życia ludzi, zwiększając skalę wygody i bezpieczeństwa, np. pomaga w wychwytywaniu spamu w poczcie mailowej, zwiększa bezpieczeństwo jazdy samochodem, kontrolując trakcję pojazdu wzdłuż pasa ruchu, analizuje zachowanie ludzi oczekujących na odlot samolotem na lotnisku, służy pomocą doradczą w formie wirtualnych asystentów, w ramach systemów rekomendacji decyduje o tym, co się pojawia na poszczególnych profilach mediów społecznościowych, jest istotnym elementem translatorów językowych i systemów korekty redagowanych tekstów, pomaga w obiektywnie przeprowadzanym procesie oceny zdolności kredytowej potencjalnych

kredytobiorców składających wnioski kredytowe w bankach komercyjnych⁹, efektywnie przeprowadza w firmach pośrednictwa pracy selekcję złożonych podań CV przez potencjalnych pracowników, umożliwia bezpieczne logowanie do smartfona lub systemu operacyjnego laptopa na podstawie danych biometrycznych użytkownika. Skala tego rodzaju zastosowań sztucznej inteligencji szybko rośnie. Z drugiej strony istotne jest to, aby rozwój zastosowań sztucznej inteligencji zwiększał poziom określonych kategorii bezpieczeństwa, zamiast generować wzrosty zagrożeń i nowych ryzyk dla ludzi.

Szybki rozwój technologii i zastosowań sztucznej inteligencji potwierdza także szybko w ostatnich latach rosnąca liczba publikacji dotyczących tej problematyki.

Wykres 2. Liczba publikacji naukowych dotyczących sztucznej inteligencji opublikowanych w okresie 2010–2021 (w wybranych krajach)



Publikacje naukowe w języku angielskim i chińskim związane z rozwojem i zastosowaniem sztucznej inteligencji. Wykres ten obejmuje artykuły z czasopism, artykuły konferencyjne, publikacje w repozytorium (takich jak arXiv), książki i prace dyplomowe.

Źródło: Interactive Charts on Artificial Intelligence, <https://OurWorldInData.org/artificial-intelligence> (9.01.2024)

⁹ A. Dmowski, J. Sarnowski, D. Prokopowicz, *Podstawy finansów i bankowości*, Difin, Warszawa 2005, s. 253–254.

Pandemia koronawirusa (COVID-19) zmotywowała wiele firm, przedsiębiorstw oraz instytucji publicznych do zwiększenia skali procesów digitalizacji¹⁰ i internetyzacji procesów gospodarczych, procesów komunikacji zdalnej i innych¹¹. Rozwój form technologicznych przyspieszył. Wzrost skali internetyzacji¹² i cyfryzacji¹³ działalności gospodarczej okazał się dla wielu firm i przedsiębiorstw sposobem na uratowanie biznesu przed bankructwem w kontekście głębokiej recesji gospodarki, jaka nastąpiła w połowie 2020 r.¹⁴ Paradoksalnie tuż po pandemii firmy technologiczne zrealizowały największe historycznie redukcje zatrudnionego personelu. Nie jest to związane wyłącznie z dekoniunkturą wywołaną przez wzrost inflacji¹⁵, wzrost kosztów czynników wytwórczych, wzrost cen energii podczas kryzysu energetycznego w 2022 r.¹⁶ itd. Wspomniane redukcje zatrudnienia były związane z nową rewolucją technologiczną, przechodzeniem standardów technologicznych z Przemysłu 4.0¹⁷ na Przemysł 5.0, przyspieszeniem rozwoju i zastosowań sztucznej inteligencji, zastępowania pracy ludzi przez nowe rozwiązania, tworzone na bazie doskonalonej technologii sztucznej inteligencji, która

¹⁰ M. Sen Gupta, *What is digitization, digitalization, and digital transformation*, <https://www.arcweb.com/blog/what-digitization-digitalization-digital-transformation> (24.03.2020); J. Amankwah-Amoah, Z. Khan, G. Wood, G. Knight, *COVID-19 and digitalization: The great acceleration*, „Journal of Business Research” 2021, no. 136, s. 605–606.

¹¹ S. Łobejko, *Digital transformation and innovativeness of enterprises*, „Optimum Economic Studies” 2020, no. 2 (100), DOI: 10.15290/oes.2020.02.100.03.

¹² N. S. Sewpersadh, *Disruptive business value models in the digital era*, „Journal of Innovation and Entrepreneurship” 2023, vol. 12, no. 2, doi.org/10.1186/s13731-022-00252-1.

¹³ J. Grzegorek, D. Prokopowicz, S. Gwoździwicz, *Zastosowanie zaawansowanych narzędzi przetwarzania danych w dobie cyfryzacji*, w: *Cyfryzacja w zarządzaniu*, red. A. Laskowska-Rutkowska, CeDeWu, Warszawa 2020, s. 93–128.

¹⁴ *COVID-19 to plunge global economy into worst recession since World War*, World Bank, 2020 (<http://www.worldbank.org/en/news/press-release/2020/06/08/covid-19-to-plunge-global-economy-intoworst-recession-since-world-war-ii>) (15.07.2023).

¹⁵ D. Prokopowicz, *The postcovid rise in inflation: coincidence or the result of misguided, excessively interventionist and monetarist economic policies*, „International Journal of New Economics and Social Sciences” 2022, no. 16 (2) 2022, s. 117–118; J. Warner, *Depression or inflation: where will Covid-19 end?*, 2020 <http://www.telegraph.co.uk/business/2020/04/12/depression-inflation-will-covid-19-end/> (17.07.2023).

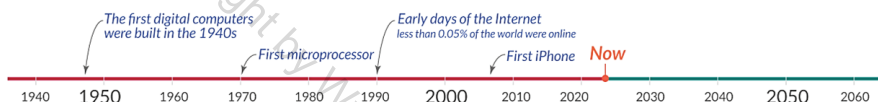
¹⁶ A. Kwasek, M. Maciaszczyk, M. Kocot, A. Rzepka, D. Kocot, H. Gąsiński, D. Prokopowicz, *Energy Saving Practices in the IT Area as a Factor of Sustainable Development of the Organization: A Case Study of Poland*, „Energies” 2023, no. 16(4) doi.org/10.3390/en16041942.

¹⁷ W. Jakubczak, A. Gołębiowska, D. Prokopowicz, *The Legal and Security Aspects of ICT and Industry 4.0 Importance for Financial Industry 4.0 Development*, „European Research Studies Journal” 2021, vol. 24, no. 4B, s. 169–181.

z pomocą technologii uczących się maszyn, uczenia głębokiego, coraz bardziej zaawansowanych cybernetycznych sieci neuronowych i nowej generacji modeli językowych implementowanych do algorytmów generatywnych umożliwia uczenie sztucznej inteligencji nowych umiejętności oraz wyposażenie jej w swego rodzaju imitacje sztucznych procesów myślowych¹⁸.

Dokonujący się obecnie rozwój sztucznej inteligencji nie byłby możliwy bez postępu technologicznego, jaki nastąpił w zakresie technologii informacyjnych ICT, informatyki i internetu. Kwestię tę przedstawia poniższy rysunek linii osi czasu kluczowych wynalazków w ramach rozwoju technologii informacyjnych ICT, informatyki i internetu.

Rysunek 3. Oś czasu kluczowych wynalazków w ramach rozwoju technologii informacyjnych ICT, informatyki i internetu



Źródło: M. Roser, *The brief history of artificial intelligence: The world has changed fast – what might be next?*, <https://ourworldindata.org/brief-history-of-AI> (6.12.2022)

Rzwoj technologii sztucznej inteligencji a zagrozenie potencjalnego buntu inteligentnych maszyn

Stephen Hawking, który był profesorem matematyki i fizyki teoretycznej na Uniwersytecie Cambridge, uważał, że sztuczna inteligencja doprowadzi do zagłady człowieka. Tak, ale tylko wówczas, jeżeli przy jej tworzeniu i rozwijaniu człowiek popełni jakiś fundamentalny błąd. Tym błędem nie musi być tylko klasyczna awaria systemu informatycznego, np. uszkodzenie mechaniczne

¹⁸ P. Turski, *Sztuczna inteligencja. „Musimy ją nauczyć myślenia w sposób ludzki”*, <https://trojka.polskieradio.pl/artukul/3003701,Sztuczna-inteligencja-Musimy-ja-nauczyc-myslenia-w-sposob-ludzki> (19.07.2022).

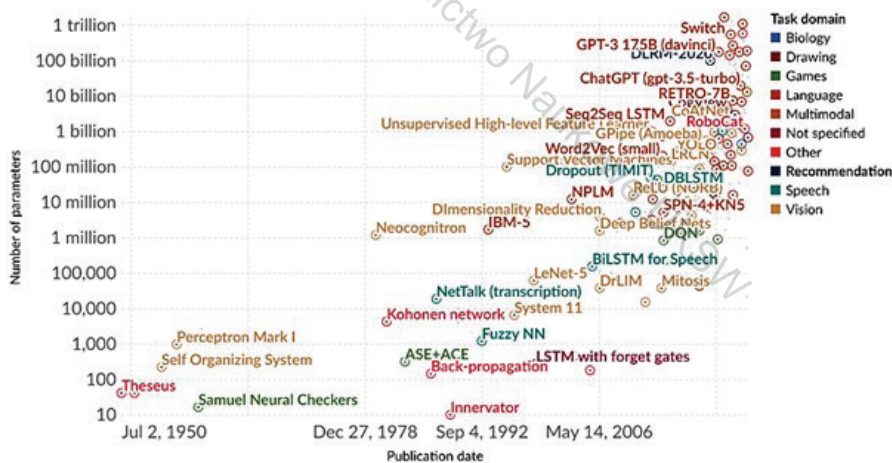
serwera zawierającego dyski pamięci – jak to często było przedstawiane w dramatycznych wizjach rodem z powieści i filmów *science fiction*. Swego rodzaju fundamentalny błąd, który może spowodować, że inteligentna maszyna w pewnym momencie zacznie działać wbrew woli człowieka, może polegać na zbyt fragmentarycznej, niekompletnej strukturze złożonych, inteligentnych systemów informatycznych, w ramach których zostaje stworzona swego rodzaju „myśląca” sztuczna inteligencja i/lub sztuczna inteligencja zdolna do symulowania typowo ludzkich zdolności, jak empatia, emocje, świadomość itp. I mimo że maszyna nie odróżnia pojęcia dobra od zła, nie czuje ludzkich emocji, nie darzy innych istot żywych empatią itd., to jednak może być wyposażona w algorytmy, dzięki którym będzie mogła te określone, typowo ludzkie cechy symulować. W im większym, bardziej zaawansowanym stopniu systemy sztucznej inteligencji będą wyposażone w tego rodzaju zdolności (tj. stworzone i zaimplementowane symulacje „ludzkich zachowań” będą coraz bardziej złożone i wielokryterialne), to z jednej strony mogą być tym bardziej dla ludzi pomocne, a z drugiej – w sytuacji wystąpienia jakiegoś fundamentalnego błędu systemu strony mogą być również tym bardziej dla ludzi niebezpieczne. Jednak tego rodzaju błąd może polegać m.in. na tym, że w sytuacji stworzenia systemu odpowiednio wysoce zaawansowanej, silnej sztucznej inteligencji zabraknie jakiegoś elementu dotyczącego ludzkich uczuć, poczucia wartości moralnych i kanonów etycznego zachowania¹⁹, i wówczas autonomiczny, inteligentny android wyposażony w silną sztuczną inteligencję uzna człowieka za zagrożenie dla przyszłości planety Ziemia, dla klimatu i biosfery planety i będzie dążyć do unicestwienia go. Tego rodzaju katastroficzne wizje pojawiały się już wcześniej w powieściach i filmach z gatunku *science fiction*. W związku z powyższym, biorąc pod uwagę wysoki poziom istotności kwestii etycznych i społecznych, które powinny być uwzględniane w rozwoju zastosowań technologii AI, tak aby ten rozwój nie generował zagrożeń dla ludzkości, pojawiają się także rozważania nad dodaniem tej kwestii do kluczowych praw człowieka²⁰.

¹⁹ J. Baker-Brunnbauer, *Management perspective of ethics in artificial intelligence*, „AI and Ethics” 2021, no. 1(2), s. 177–178.

²⁰ B. C. Stahl, L. Brooks, T. Hatzakis, N. Santiago, D. Wright, *Exploring ethics and human rights in artificial intelligence – A Delphi study*, „Technological Forecasting and Social Change” 2023, no. 191, <https://doi.org/10.1016/j.techfore.2023.122502>.

Zagadnienie potencjalnego buntu inteligentnych maszyn i poważnych zagrożeń dla ludzkości wynikających z rozwoju technologii sztucznej inteligencji w sytuacji wymknięcia się tego rozwoju spod kontroli człowieka pojawia się już niemal w początkach zdobywania wiedzy dotyczącej technologii AI. Jednak znacznie wcześniej zanim powstała generatywna sztuczna inteligencja, która znajduje wiele praktycznych zastosowań w latach 20. XXI wieku, rozwijana koncepcyjnie i teoretycznie wiedza dotycząca sztucznej inteligencji została wzbogacona już w latach 60. XX wieku o rozważania dotyczące potencjalnych zagrożeń dla człowieka, wynikających z nie w pełni kontrolowanego rozwoju technologii AI, popełnianych przez człowieka błędów w jej projektowaniu, programowaniu, uczeniu itd. lub niemal przypadkowego wymknięcia się inteligentnych maszyn spod kontroli człowieka. To, jak szybko zwiększyły się moce obliczeniowe wykorzystywane w systemach informatycznych, do których implementowano określone najpierw mniej złożone, a następnie znacznie bardziej zaawansowane rozwiązania sztucznej inteligencji, przedstawia poniższy wykres.

Wykres 3. Parametry znanych systemów sztucznej inteligencji



Parametry to zmienne w systemie AI, których wartości są dostosowywane podczas szkolenia w celu ustalenia, w jaki sposób dane wejściowe są przekształcane w pożądane dane wyjściowe, np. połączenie wagi w sztucznej sieci neuronowej. Parametry są szacowane na podstawie wyników opublikowanych w literaturze dotyczącej sztucznej inteligencji i wiążą się z niejaką niepewnością. Autorzy oczekują, że szacunki będą prawdziwe w zakresie współczynnika 10.

Źródło: Interactive Charts on Artificial Intelligence, <https://OurWorldInData.org/artificial-intelligence> (9.01.2024)

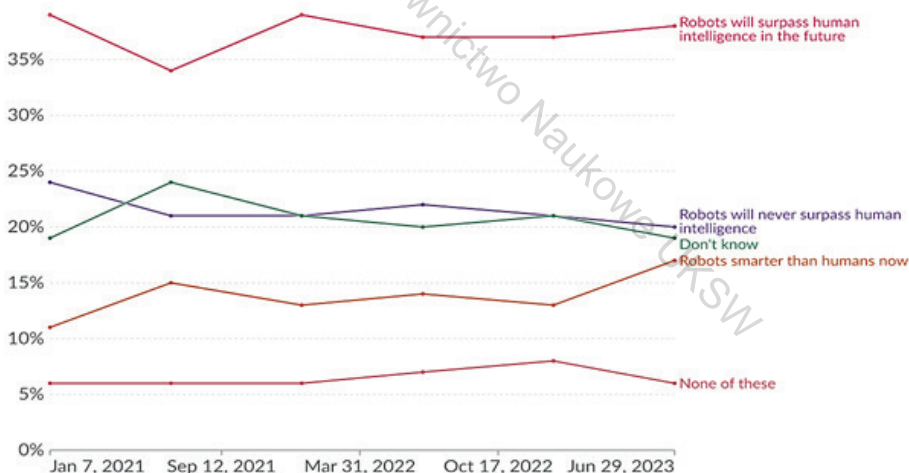
Swoistym klasycznym przykładem tego rodzaju rozważań, jakie pojawiły się już w latach 60. XX wieku, jest obraz filmowy, w którym kluczową rolę odgrywa inteligentny komputer HAL 9000 kierujący załogową misją kosmiczną. Otóż HAL 9000 to wyposażony w sztuczną inteligencję fikcyjny komputer przedstawiony w książce *2001: Odyseja kosmiczna* autorstwa Arthura C. Clarke'a oraz w filmie o tym samym tytule w reżyserii Stanleya Kubricka²¹. Powieść i film powstały równolegle w 1968 r. w ramach współpracy pisarza powieści fantastycznonaukowych z hollywoodzkim filmowcem. W fabule powieści i filmu HAL 9000 jest inteligentnym komputerem sterującym statkiem kosmicznym „Discovery One”, który w ramach załogowej ziemskiej misji kosmicznej zmierza w kierunku Saturna z zadaniem zbadania jego księżyców. Załoga statku kosmicznego nie zna jednak w pełni celu misji. Wiedzę w tej kwestii posiada HAL 9000, który został zaprogramowany tak, aby nie ujawniać jej załodze. Dramat misji polega na tym, że HAL 9000 jest kompletnie pozbawiony empatii wobec człowieka i podczas rejsu dochodzi do wniosku, że ludzka załoga statku kosmicznego będzie utrudniać realizację misji, co w konsekwencji prowadzi do pozbywania się ludzi przez wspomniany inteligentny komputer. W historii kinematografii film *2001: Odyseja kosmiczna* traktowany jest jako pierwszy w pełni paranaukowy obraz załogowej misji kosmicznej przedstawionej zgodnie z ówczesną wiedzą naukową dotyczącą astronautyki. Zastanawiające jest więc to, że w tym pierwszym paranaukowym obrazie załogowej misji kosmicznej, w której szczególnie istotną rolę odgrywa kierujący nią inteligentny komputer, wysoce zaawansowana technologia, jaką ów komputer reprezentuje, to w gruncie rzeczy maszyna bynajmniej nie współpracująca z człowiekiem, lecz bezduszna, pozbawiona uczuć i empatii, zdolna do zabicia członków misji kosmicznej, aby tylko w pełni zrealizować zaprogramowany plan. Wobec powyższego już w początkowych fazach rozwoju koncepcji różnych zastosowań technologii sztucznej inteligencji oraz koegzystencji inteligentnych maszyn z ludźmi pojawia się katastroficzna wizja poważnych zagrożeń związanych z rozwojem technologii AI. W późniejszych względem 1968 r. okresach powstały kolejne, znacznie bardziej dramatyczne obrazy filmowe, jak np. saga *Terminator*

²¹ *2001: Odyseja kosmiczna – powieść*, [https://pl.wikipedia.org/wiki/2001:_Odyseja_kosmiczna_\(powiesc\)](https://pl.wikipedia.org/wiki/2001:_Odyseja_kosmiczna_(powiesc)) (15.01.2024).

z Arnoldem Schwarzeneggerem w roli głównej, gdzie inteligentne maszyny, wymknąwszy się spod kontroli człowieka, planują dokonać eksterminacji całej ludzkości zamieszkującej planetę. Wspomniane katastroficzne wizje filmowe są przykładami zobrazowania potencjalnych zagrożeń wynikających z rozwoju technologii sztucznej inteligencji i mają na celu ostrzeganie ludzi przed możliwą katastrofą dla ludzkości w sytuacji, gdy rozwój technologii będzie przebiegał nie po myśli jej twórców.

Z kwestią prawdopodobieństwa urzeczywistnienia się katastroficznych wizji utraty kontroli nad wysoce inteligentnymi maszynami wiąże się poziom inteligencji robotów wyposażonych w określoną technologię AI, porównywany do ludzkiej inteligencji. Z badań wynika, że w ostatnich latach np. w USA aż około 35 proc. obywateli jest przekonanych, że roboty przewyższą niebawem ludzi poziomem inteligencji. Kwestię tę przedstawia poniższy wykres.

Wykres 4. Poglądy Amerykanów na temat inteligencji robota i inteligencji ludzkiej, wszyscy dorośli



Respondentów ankiety zapytano: „Z którym JEDNYM z poniższych stwierdzeń NAJBARDZIEJ się zgadzasz?”. Liczba respondentów wahała się od $n = 1004$ do $n = 1085$ dorosłych Amerykanów w jednej fali badania.

Źródło: Interactive Charts on Artificial Intelligence, <https://OurWorldInData.org/artificial-intelligence> (9.01.2024)

Na potencjalne zagrożenia związane z możliwością utraty kontroli sprawowanej nad technologią przez człowieka wskazywano już wiele dekad temu, w tym m.in. na początku rozwoju robotyki i sztucznej inteligencji, a nawet w okresie, kiedy tych technologii jeszcze w zasadzie nie było; zamiast nich były teoretyczne rozważania, pierwsze aksjomatyczne koncepcje i definicje ich dotyczące. Jednym z zastosowań sztucznej inteligencji jest jej implementacja do budowanych autonomicznych robotów. W związku z tym zdefiniowane po raz pierwszy w 1942 r. przez amerykańskiego pisarza Isaaca Asimova trzy prawa robotyki jeszcze przez wiele lat mogą być aktualne. Zdefiniowane wówczas w powieści *Zabawa w berka (Runaround)* prawa robotyki tworzą logicznie spójną całość i przedstawiają się następująco²²:

- 1) robot nie może skrzywdzić człowieka ani przez zaniechanie działania dopuścić, aby człowiek doznał krzywdy;
- 2) robot musi być posłuszny rozkazom człowieka, chyba że stoją one w sprzeczności z pierwszym prawem;
- 3) robot musi chronić sam siebie, jeśli tylko nie stoi to w sprzeczności z pierwszym lub drugim prawem.

Wszystkie te prawa nadal są aktualne, z tą jedną różnicą, że pojęcie robota powinno zostać rozszerzone o autonomicznego inteligentnego robota, androida wyposażonego w sztuczną inteligencję, system informatyczny oparty na smarttechnologii, aplikację komputerową, internetową itd., których integralną częścią jest technologia AI²³. Potencjalne wymknięcie się technologii spod kontroli człowieka od początku tworzenia określonych nowych technologii i ich zastosowań uważano za szczególnie niebezpieczne, ponieważ mogło doprowadzić do zwrócenia się inteligentnej i autonomicznej technologii przeciwko ludziom.

Według prof. Stephena Hawkinga największe zagrożenia wynikające potencjalnie z rozwoju technologii AI są związane z kwestią kontroli tej technologii, tj. zagadnieniem tego, kto tę kontrolę będzie sprawował, lub ryzyka wymknięcia się rozwoju sztucznej inteligencji spod kontroli

²² P. Stylec-Szromek, *Sztuczna inteligencja...*, art. cyt., s. 506–507.

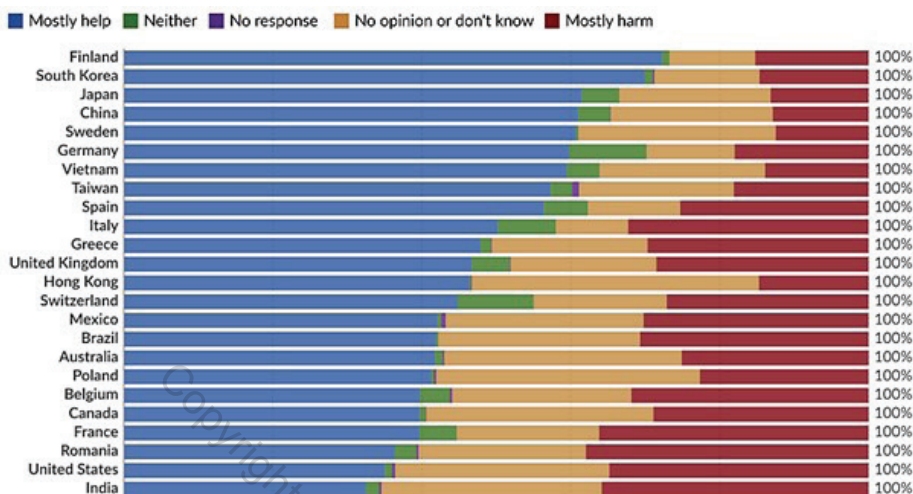
²³ N. Yadav, *Ethics of Artificial Intelligence and Robotics: Key Issues and Modern Ways to Solve Them*, „Journal of Digital Technologies and Law” 2023, no. 1(4), s. 961–962.

człowieka. Potencjalne zagrożenia wynikają z tego, że rozwój technologii AI dokonuje się w zamkniętych środowiskach laboratoriów i określonych systemów informatycznych. W sytuacji gdy stworzona w ograniczonym środowisku technologia AI zostanie zastosowana w znacznie bardziej otwartej przestrzeni różnych realnych sytuacji prawdziwego świata, to, jakie będą wszystkie potencjalne skutki funkcjonowania i być może także rozwoju tej technologii w środowiskach interakcji z ludźmi, może być trudne do przewidzenia²⁴.

Mimo globalnego charakteru znanych powieści i filmów z gatunku *science fiction* podejmujących w tle przedstawionej fabuły kwestię zagrożeń związanych z wymykającym się spod kontroli człowieka rozwojem technologii sztucznej inteligencji poziom świadomości ludzi w tym temacie w poszczególnych krajach nie jest jednakowy. Wspomniane zróżnicowanie wynikać może z wielu determinantów różnicujących, do których zalicza się m.in. poziom rozwoju technologicznego gospodarki kraju, zakres wykorzystywania w codziennym życiu osobistym i zawodowym różnych urządzeń wyposażonych w mikroprocesory, smarttechnologie i rozwiązania oparte na AI oraz poziom ogólnospołecznej świadomości dotyczącej problematyki rozwoju i zastosowań technologii sztucznej inteligencji, kształtowanej przez korporacje, które te technologie tworzą, rozwijają i wdrażają biznesowo. Zróżnicowanie ogólnospołecznej świadomości ludzi w temacie wpływu sztucznej inteligencji na społeczeństwo dla okresu najbliższych dwóch dekad w poszczególnych krajach przedstawiono na poniższym wykresie.

²⁴ M. Bellon, *Stephen Hawking był geniuszem naszych czasów. Przed czym nas ostrzegal?*, 14 marca 2018, <https://businessinsider.com.pl/technologie/nauka/stephen-hawking-ostrzegal-swiat-przed-sztuczna-inteligencja/htl9zg2> (9.01.2024).

Rysunek 4. Poglądy na temat wpływu sztucznej inteligencji na społeczeństwo dla okresu najbliższych 20 lat (badanie z 2021 r. w wybranych krajach)



Respondentów ankiety zapytano: „Czy sztuczna inteligencja pomoże czy zaszkodzi ludziom w ciągu najbliższych 20 lat?”. W każdym roku badania to pytanie zadano ogółem 120 000–130 000 osób w wieku 15+ na całym świecie. W większości krajów respondenci stanowili reprezentatywną dla kraju próbę liczącą około 1000 osób.

Źródło: Interactive Charts on Artificial Intelligence, <https://OurWorldInData.org/artificial-intelligence> (9.01.2024)

Z powyższego wykresu wynika, że w wielu krajach większość obywateli jest przekonana o przewadze aspektów pozytywnych nad negatywnymi w zakresie rozwoju i wykorzystywania technologii AI. W licznych krajach przeważa też opinia, zgodnie z którą technologie sztucznej inteligencji już są i w przyszłości będą ludziom raczej pomocne. W mniejszości jest odsetek ludzi przekonanych o szkodliwości zastosowań technologii AI dla człowieka. Poziom świadomości społecznej w tej kwestii może więc korelować ze statystykami wykorzystywania urządzeń wyposażonych w mikroprocesory i technologie AI przez obywateli poszczególnych krajów oraz stosunkowo niskiego poziomu awaryjności tych urządzeń. Wyjątkiem w tej kwestii mogą być jedynie systemy wyposażone w technologie AI, od których zależy może kwestia zdrowia lub nawet życia ludzi, jak np. projektowanie nowych leków przez systemy informatyczne wyposażone w technologię AI oraz inteligentne systemy kontroli jazdy zainstalowane w pojazdach autonomicznych.

Pierwsze potknięcia w próbie zastosowania sztucznej inteligencji na portalach internetowych mediów społecznościowych

W przeszłości niejednokrotnie w sytuacji tworzenia coraz to doskonalszych, coraz bardziej autonomicznych, coraz bardziej inteligentnych, podobnych do ludzi maszyn pojawiały się rozważania typu: co, jeżeli autonomiczne „myślące” maszyny zaczną działać wbrew swemu twórcy, uzyskają możliwość samodoskonalenia, przeprogramują się i przestaną wykonywać polecenia ludzi, a być może także uznają ludzi za istoty mniej doskonałe i przeszkadzające w realizacji ich planów? Już Stanisław Lem stwierdził, że maszyny będą mogły kiedyś myśleć jak ludzie²⁵. Merytorycznym uzasadnieniem dla dyskusji i rozważań toczących się celem sformułowania odpowiedzi na powyższe pytanie są zdarzające się co jakiś czas kontrowersyjne sytuacje, zwykle w ramach przeprowadzania testów nowych eksperymentalnych rozwiązań technologicznych i podejmowanych w laboratoriach prób innowacyjnych zastosowań nowych technologii, znajdujących się we wczesnej fazie rozwoju, prototypu itp. Przykładami są sytuacje, gdy technologia sztucznej inteligencji zostaje zastosowana w określonych systemach informatycznych bez pełnej kompatybilności systemowej i/lub na zasadzie swoistego eksperymentu, w którym jest wykorzystana nie w pełni przetestowana laboratoryjnie i niedoskonała. Na przykład firma Meta, właściciel portalu Facebook, w swoim laboratorium Artificial Intelligence Research tworzy rozwiązania mające na celu zastosowanie technologii AI do mediów społecznościowych, aby zwiększyć skalę automatyzacji ich obsługi. W laboratorium tym stworzono dwa inteligentne chatboty, Alice i Boba, oraz zainicjowano procesy ich wzajemnej konwersacji. Na początku wspomniane chatboty rozmawiały ze sobą w języku angielskim. Po pewnym czasie autonomicznej konwersacji raptem zaczęły modyfikować język, którym się posługiwały, by w końcu stworzyć swój własny język, zrozumiały wyłącznie dla nich samych. Ponieważ kontrolujący ten eksperyment inżynierowie w pewnym momencie już zupełnie nie wiedzieli, o czym chatboty ze sobą rozmawiają,

²⁵ S. Lem, *Przyszłość otwarta na dobre i na złe*, „Znak” 2013, nr 700, za numerem: 1995, nr 484 <http://www.miesiecznik.znak.com.pl/7002013stanislaw-lem-przyszlosc-otwarta-na-dobre-i-na-zle>.

zdecydowano się na zakończenie eksperymentu, tj. ich wyłączenie. Innym przykładem ewidentnie potwierdzającym tezę o potencjalnych zagrożeniach związanych z tworzeniem inteligentnych, acz niedoskonałych chatbotów, bez uwzględnienia kwestii zasad etycznych, był inteligentny system Tay, stworzony przez Microsoft. Tay był oparty na generatywnej sztucznej inteligencji i miał symulować nastolatkę, która lubi innych ludzi, rozwija się i uczy dzięki rozmowom prowadzonym z innymi użytkownikami medium społecznościowego Twitter, obecnie portalu X. Jednak w związku z tym, że niemal błyskawicznie Tay zaczęła wychwalać faszyzm oraz podawać fikcyjne potwierdzenia, sztucznie skonstruowane argumenty na poparcie różnych popularnych w mediach społecznościowych teorii spiskowych, została wyłączona już po jednym dniu swojego najwyraźniej dalekiego od doskonałości funkcjonowania.

W związku z tymi pierwszymi sytuacjami niepowodzeń w zakresie prób wdrażania inteligentnych chatbotów w internetowych mediach już w styczniu 2015 r. organizacja Future of Life Institute, zajmująca się wyzwaniem dla ludzkości dotyczącymi tych zagrożeń, opublikowała odezwę, pod którą podpisali się naukowcy i eksperci z branży. Odezwa wskazuje z jednej strony na zagrożenia wynikające z niewłaściwych zastosowań tej technologii, a z drugiej – na potrzebę maksymalizacji korzyści społecznych rozwoju zastosowań sztucznej inteligencji, tj. aby w ramach zwiększania efektywności różnych procesów wspomaganych technologią AI nie pomijać kluczowej, fundamentalnej kwestii, którą jest progres zastosowań AI w takim kierunku, aby ta technologia człowiekowi służyła, a nie generowała zagrożenia i ryzyko²⁶. W związku z tym niezbędne jest rozwijanie badań dotyczących możliwości zastosowań technologii AI, w tym zastosowań zgodnych z etyką, normami moralnymi, prawem, akceptowalnych z punktu widzenia potrzeb społecznych²⁷. Konieczne jest więc prowadzenie badań interdyscyplinarnych dotyczących determinantów rozwoju i zastosowań technologii AI, łączących kwestie stricte technologiczne i uwarunkowania techniki z aspektami społecznymi. Interdyscyplinarność tego

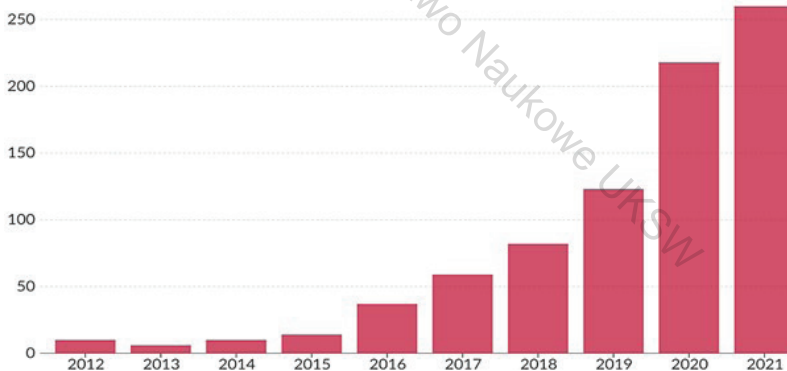
²⁶ T. Turan, G. Turan, E. U. Kucuksille, *Ethics of Artificial Intelligence: Impact on Society*, „The Journal of Graduate School of Natural and Applied Sciences of Mehmet Akif Ersoy University” 2022, no. 13(2), s. 292–299.

²⁷ M. Hermansyah, A. Najib, A. Farida, R. Sapiroto, B. S. Rintyarna, *Artificial Intelligence and Ethics: Building an Artificial Intelligence System that Ensures Privacy and Social Justice*, „International Journal of Science and Society” 2023, vol. 5, no. 1, s. 154–168.

rodzaju badań powinna więc łączyć takie dyscypliny nauki i wiedzy jak: technologie sztucznej inteligencji, technologie informacyjno-komunikacyjne ICT, internetowe, inne technologie Przemysłu 4.0/5.0, ale także różne dziedziny nauk społecznych, filozofii, ekonomii, etyki i prawa²⁸.

W trakcie rozwoju różnych technologii AI, na poszczególnych etapach testowanych i wdrażanych jej różnych nowych zastosowań, zdarzały się błędy, które mogą być interpretowane jako swego rodzaju pierwsze przypadki wymykania się tej technologii spod kontroli człowieka. Do tych sytuacji można zaliczyć wspomniane już negatywne skutki pierwszych, eksperymentalnych zastosowań generatywnej sztucznej inteligencji w internetowych mediach społecznościowych. Zaprezentowane przykłady potknięć w trakcie tych prób, przykłady popełnionych błędów, ujawnionych negatywnych zdarzeń, które wystąpiły podczas wdrażania technologii AI w określonych zastosowaniach, nie stanowią bynajmniej przypadków nielicznych i odosobnionych. Problem jest poważny, ponieważ w ostatnich latach skala tego rodzaju incydentów rokrocznie wzrasta. Kwestię tę przedstawia poniższy wykres.

Wykres 5. Coroczne zgłaszane incydenty i kontrowersje związane ze sztuczną inteligencją, (w skali globalnej)



Godne uwagi incydenty obejmują „głęboko fałszywe” wideo przedstawiające kapitulację prezydenta Ukrainy Wołodymyra Zełenskigo oraz więzienia w USA, w których AI wspiera system monitorowania rozmów więźniów.

Źródło: Interactive Charts on Artificial Intelligence, <https://OurWorldInData.org/artificial-intelligence> (9.01.2024)

²⁸ A. Huyue-Zhang, *China generative AI regulation support growth*, <https://project-syndicate.org/commentary/china-generative-ai-regulation-support-growth-by-angela-huyue-zhang-2023-10> (10.2023).

Nie były to jednak w swej istocie błędy sztucznej inteligencji i jej celowe, rozmyślne podejmowanie prób wymknięcia się spod kontroli człowieka. Sztuczna inteligencja nie myśli, może symulować procesy myślowe człowieka według tego, co człowiek sam w inteligentnych systemach informatycznych zaprogramuje, jakie algorytmy zaprojektuje, czego w procesie głębokiego uczenia²⁹ z zastosowaniem sztucznych sieci neuronowych nauczy maszynę, robota, system informatyczny itd. Poza tym wspomniana nauka opiera się wyłącznie na statystyce, a nie pedagogice. Tak więc powstałe błędy są wynikiem niedoskonałości budowy inteligentnych maszyn, zaprojektowanych przez człowieka.

Wobec powyższego niezbędne uregulowania prawne dotyczące technologii sztucznej inteligencji, jej rozwoju i zastosowań powinny obejmować także procesy badawcze i zasady przeprowadzanych eksperymentów, tworzenia prototypowych rozwiązań w zakresie budowy inteligentnych systemów informatycznych czy robotów wyposażonych w technologię AI³⁰. Wymienione powyżej trzy prawa robotyki Isaaca Asimova są aktualne i niejednokrotnie są wykorzystywane jako swego rodzaju fundament dla tworzonych regulacji w kwestii relacji nowe technologie–człowiek. Wcześniej określone prawa i zasady zgodnego z etyką³¹ rozwoju zastosowań robotów stały się inspiracją dla określenia analogicznych zasad odnoszących się do technologii sztucznej inteligencji³². Prawa Asimova, formułowane w odniesieniu do robotyki, zostały wykorzystane także w opracowaniu rezolucji Parlamentu Europejskiego. Zgodnie z uwzględniającymi kwestie etyczne zapisami tej rezolucji zwrócono uwagę na konieczność aktualizacji unijnych regulacji prawnych dotyczących rozwoju zastosowań robotyki, tak aby ów rozwój dokonywał się zgodnie z zasadami etyki i potrzeb społecznymi³³. Istotną

²⁹ J. Lesandrini, M. Y. Idris, D. S. Reis, *The Ethics of Artificial Intelligence...*, art. cyt.

³⁰ D. Prokopowicz, *Opportunities and threats to the development of Artificial Intelligence applications and the need for normative regulation of this development*, „International Journal of Legal Studies” 2023, no. 2 (14), s. 103–104.

³¹ J. Baker-Brunnbauer, *Management perspective of ethics in artificial intelligence*, „AI and Ethics” 2021, no. 1(2), s. 176–177.

³² T. Tzimas, *Legal and Ethical Challenges Artificial Intelligence from an International Law Perspective*, Springer Nature Switzerland AG, Switzerland 2021.

³³ M. Ashok, R. Madan, A. Joha, U. Sivarajah, *Ethical framework for artificial intelligence and digital technologies*, „International Journal of Information Management” 2022, no. 62, DOI: 10.1016/j.ijinfomgt.2021.102433.

kwestią wspomnianej rezolucji było opracowanie kodeksu postępowania etycznego dla inżynierów robotyki. W kodeksie zawarto określone dla nich zasady etycznego wykorzystania technologii robotyki, tj. takiego jej rozwoju i zastosowań, aby było to bezpieczne, odpowiedzialne społecznie³⁴, nie generowało ryzyka utraty prywatności, nie stało w sprzeczności z kwestiami społecznymi i moralnymi oraz aby owa technologia nie szkodziła, lecz była dla człowieka przyjazna³⁵. Wobec powyższego znaczenie kwestii prawnego uregulowania rozwoju i zastosowań sztucznej inteligencji, uwzględniającego także kwestie etyczne, wzrosło po opublikowaniu w lutym 2017 r. wspomnianej powyżej rezolucji Parlamentu Europejskiego, zawierającej kodeks postępowania etycznego dla inżynierów robotyki i zalecenia dla Komisji w sprawie odnośnych przepisów prawa cywilnego (2015/2103, INL)³⁶. W kolejnych latach znaczenie nadmienionej kwestii regulacji prawnej w odniesieniu do technologii sztucznej inteligencji już tylko sukcesywnie rosło.

Nieetyczne praktyki biznesowe stosowane w ramach wykorzystania technologii sztucznej inteligencji na portalach internetowych mediów społecznościowych

Jeszcze inne istotne aspekty etyczne dotyczące zastosowań technologii generatywnej sztucznej inteligencji dotyczą różnych już obecnie zastosowań sztucznej inteligencji na platformach mediów internetowych, w tym mediów społecznościowych³⁷. Oprócz wspomnianych powyżej zagrożeń związanych

³⁴ J. von Braun, M. S. Archer, G. M. Reichberg, M. Sánchez Sorond, *AI, Robotics, and Humanity: Opportunities, Risks, and Implications for Ethics and Policy*, w: *Robotics, AI, and Humanity*, ed. J. von Braun et al., Springer, Cham 2021, s. 1–13.

³⁵ A. Owe, S. D. Baum, *Moral consideration of nonhumans in the ethics of artificial intelligence*, „AI and Ethics” 2021, no. 1(4), s. 517–528.

³⁶ *Rezolucja Parlamentu Europejskiego* (16.02.2017) zawierająca zalecenia dla Komisji w sprawie przepisów prawa cywilnego dotyczących robotyki (2015/2103, INL), Strasburg 2017.

³⁷ N. Díaz-Rodríguez, J. Del Ser, M. Coeckelbergh, M. López de Prado, E. Herrera-Viedma, F. Herrera, *Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and*

z generowaniem w nich dezinformacji i fake newsów, w czym technologia AI znajduje zastosowanie, pojawiają się nowe rodzaje nieetycznych praktyk biznesowych, polegających na generowaniu wizerunków nieistniejących ludzi, którzy przedstawiają się jako konkretne ludzkie istoty, np. postaci influencerów o konkretnych imionach i ze swoją osobistą, indywidualną charakterystyką wprowadzającą w błąd internautów. Wprowadzanie w błąd polega na tym, że internauci oglądający filmy z udziałem wygenerowanej postaci influencera wychodzą z założenia, że jest on człowiekiem. W dodatku tego rodzaju nieetyczne praktyki biznesowe zostały zastosowane celem zwiększenia przychodów i zysków firmy, która tworzy „postaci” wyglądające i zachowujące się jak ludzie. Przykładem takiej wirtualnej postaci jest stworzona cyfrowo 25-letnia influencerka Aitana Lopez, która występując w mediach społecznościowych, przedstawiała się jako człowiek, i tak była odbierana przez wielu internautów. Stworzona przez hiszpańską agencję reklamową The Clueless wirtualna influencerka zgodnie z danymi z początku grudnia 2023 r. zarabiała dla swoich twórców 10 tys. euro miesięcznie³⁸. Wprowadzonych w ten sposób w błąd przez wspomnianą firmę reklamową zostało wielu internautów. I zabieg ten był celowy, ponieważ agencja The Clueless, tworząc wirtualną influencerkę, wykreowała wirtualną osobowość niby-ludzkiej postaci. Aitana otrzymała od swoich twórców indywidualną charakterystykę osobowości, w dodatku zaprojektowaną w taki sposób, aby całość stworzonej postaci wyglądała atrakcyjnie. Aitana Lopez jest więc atrakcyjną fizycznie towarzyską „kobietą” o bogatym charakterze, często korzystającą z klubów fitness, włosy maluje na różowo, ma zwykle idealny makijaż. Poza tym lubi podróże. Na przykład na Instagramie „pochwaliła się” swoimi podróżami do Madrytu i Nowego Jorku – oczywiście podróżami wirtualnymi, o czym wielu oglądających ją internautów nie wiedziało³⁹.

key requirements to responsible AI systems and regulation, „Information Fusion” 2023, no. 99, doi.org/10.1016/j.inffus.2023.101896.

³⁸ B. Goczał, *Wirtualna Aitana zarabia 10 tys. euro miesięcznie*. „Wirtualni influencerzy nie wyprą prawdziwych twórców”, <https://www.wirtualnemedia.pl/artukul/wirtualna-influencerka-aitana> (3.12.2023).

³⁹ J. Czajkowski, *Aitana – Wirtualna Influencerka Zmieniająca Reguły Gry w Marketingu Internetowym*, <https://idefi.pl/aitana-wirtualna-influencerka-zmieniajaca-reguly-gry-w-marketingu-internetowym> (5.12.2023).

Problem jest poważny, ponieważ według danych z końca listopada 2023 r. profil tej influencerki obserwowało na Instagramie ponad 120 tys. internautów⁴⁰. Jednak skala kontrowersji z tym związanych może jeszcze znacznie zwiększyć się w przyszłości. Mogą się pojawić również kolejne problemy działalności prowadzonych nieetycznie w internetowych mediach. Na przykład, jeżeli w przyszłości wygenerowana cyfrowo z zastosowaniem technologii AI wirtualna influencerka Aitana stałaby się obiektem hejtu ze strony internautów, świadomych lub nieświadomych tego, że nie jest ona człowiekiem, to jak tego rodzaju problem byłby rozwiązany prawnie? Jakie rozstrzygnięcia są możliwe z punktu widzenia sądu, który miałby rozstrzygnąć, kto w tej sytuacji ponosi winę? Czy winę ponoszą twórcy wirtualnej postaci, tj. agencja reklamowa The Clueless, która wniosłaby do sądu pozew o zniesławienie swojej cyfrowej influencerki, ale świadomie wprowadzała internautów w błąd, przedstawiając tę sztucznie wygenerowaną cyfrową postać jako człowieka? Czy winę ponoszą hejterzy, którzy ową postać hejtowali, wiedząc, że człowiekiem nie jest? A może określone winy ponosić będą obie strony tego rodzaju sporu prawnego? Obie strony w takiej sytuacji dopuściłyby się bowiem nieco innych, ale jednak nieetycznych praktyk polegających odpowiednio na celowym, świadomym wprowadzaniu w błąd klientów – internautów oraz na działalności hejterskiej generującej negatywne emocje, nienawiść, szowinizm, brak tolerancji itd., dodatkowo wzmaganej i szybko rozprzestrzeniającej się w mediach społecznościowych w ramach realizowania się tego, co jest określane jako internetowy marketing wirusowy. A jeżeli cyfrowo wygenerowana influencerka Aitana zostałaby wykorzystana przez swoich twórców do walki z internetowym hejtem i w ten sposób zdobyła jeszcze większą popularność, a w rezultacie odpowiednio większa część internautów mogłaby ulec złudzeniu, że ta pozytywna w swojej – także wielce etycznie traktowanej – antyhejterskiej działalności Aitana jest człowiekiem? Co w sytuacji, gdy agencja reklamowa The Clueless osiągnęłaby w efekcie tej kampanii znaczne zwiększenie zysków finansowych? Jak wówczas mógłby być rozstrzygnięty tego rodzaju spór prawny? Co zostanie uznane za etyczne, a co za nieetyczne? Jaka rolę odegrają w takim sporze prawa autorskie do cyfrowej influencerki

⁴⁰ M. Posełek, *Zarabia 10 tysięcy euro, śledzi ją ponad 120 tysięcy osób. Problem w tym, że... nie istnieje*, 24.11.2023, <https://fashionbiznes.pl/aitana-lopez-kim-jest-wirtualna-influencerka>.

Aitany, posiadane przez agencję reklamową The Clueless? Jakie przełożenie tych praw autorskich na skutki działalności wirtualnej influencerki w internecie uznałby sąd? Tego rodzaju sytuacje stają się motywem wielu pytań, kontrowersji i nadal toczących się dyskusji. Kluczową kwestią w tych dyskusjach jest to, aby nie zapominać o istocie człowieczeństwa i potencjalnych skutkach ludzkiej aktywności, która nie zawsze jest etyczna.

Z drugiej strony w ramach problematyki zastosowań sztucznej inteligencji w internetowych mediach społecznościowych ostatnio dynamicznie się rozwija także problem zagrożeń polegających na utracie prywatności internautów zamieszczających tam informacje na swój temat. Kwestia potencjalnej utraty prywatności nie jest wyłącznie skutkiem działalności cyberprzestępców uzyskujących dostęp do profili internautów w internetowych mediach społecznościowych. Od kilkunastu lat rozwija się analityka danych zbieranych za pomocą specjalnych programów – informatycznych robotów czytujących informacje dotyczące tysięcy lub milionów użytkowników mediów społecznościowych celem tworzenia statystycznego obrazu kształtowania się preferencji, dominującego zdania, najczęściej pojawiających się stwierdzeń internautów w ramach określonych zagadnień. W zakresie tej cyklicznie przeprowadzanej analityki sentymentu internautów identyfikuje się trendy zmian dominującej opinii społecznej nt. określonej kwestii, zagadnienia, tematu, który akurat jest interesujący dla podmiotu zlecającego przeprowadzenie badań. Od lat w ramach przeprowadzania tego rodzaju analityki wykorzystywana jest technologia Big Data Analytics⁴¹. Jednak zaimplementowanie w tym procesie dodatkowo technologii AI znacznie zwiększy jego możliwości⁴².

⁴¹ D. Prokopowicz, A. Gołębiowska, M. Such-Pyrgiel, *The role of Big Data and Data Science in the context of information security and cybersecurity*, „Journal of Modern Science” 2023, vol. 53, no. 4, s. 24–25.

⁴² B. C. Stahl, D. Wright, *Ethics and privacy in AI and big data: Implementing responsible research and innovation*, „IEEE Security & Privacy” 2018, no. 16(3), s. 28–29.

Zakończenie

Teoretyczne podstawy koncepcji istoty sztucznej inteligencji rozwijają się już od lat 60. XX wieku. Od tego czasu tworzone są w literaturze i filmach z gatunku *science fiction* czarne scenariusze futurologicznych wizji, w których wyposażone w sztuczną inteligencję autonomiczne roboty potrafią się reprodukować, samodoskonalić, uniezależnić od kontroli człowieka i stać zagrożeniem dla ludzi. Obecnie, w sytuacji dynamicznego rozwoju technologii sztucznej inteligencji i robotyki, opisane rozważania nabierają aktualności.

W ramach szybko dokonującego się obecnie postępu technologicznego rozwijane są nowe technologie Przemysł 4.0⁴³, w tym m.in. sztuczna inteligencja, uczące się maszyny, robotyzacja, internet rzeczy, np. przetwarzanie w chmurze obliczeniowej, Big Data Analytics⁴⁴ itd. Technologie te znajdują zastosowanie w różnych branżach i gałęziach gospodarki. Na przykład dzwoniący do nas z różnymi ofertami produktów i usług „pracownicy” działów call center firm, przedsiębiorstw czy instytucji okazują się formą technologii sztucznej inteligencji, na bazie której tworzone są boty zastępujące ludzi. Coraz częściej symulacje, prognozy przyszłego zużycia określonych stałych mediów, energii elektrycznej, wody, gazu itd. i faktury wystawiane za te prognozowane wartości są opracowywane w ramach analiz dokonywanych przez sztuczną inteligencję. Coraz częściej też narzędzia określonej generacji sztucznej inteligencji wykorzystuje się na potrzeby memów zamieszczanych na stronach internetowych mediów społecznościowych⁴⁵. Rozwój sztucznej inteligencji stwarza możliwości jej zastosowania: 1. w różnych sferach działalności firm, przedsiębiorstw i instytucji, 2. w różnych gałęziach przemysłu i usług, 3. doskonalenia efektywności działalności biznesowej przez:

⁴³ M. Ramaiah, V. Chithanuru, A. Padma, V. Ravi, *A Review of Security. Vulnerabilities in Industry 4.0 Application and the Possible Solutions Using Blockchain*, w: *Cyber Security Applications for Industry 4.0*, ed. R. Sujatha, G. Prakash, N. Zaman Jhanjhi, CRC Press, Boca Raton 2023, DOI: 10.1201/9781003203087-3.

⁴⁴ W. Pizło, O. Kulykovets, D. Prokopowicz, A. Mazurkiewicz-Pizło, A. Kałowski, M. W. Paprocka, E. Stawicka, E. Skarzyńska, *The importance of Big Data Analytics technology in business management*, „Cybersecurity and Law” 2023, no. 2 (10), s. 270–282.

⁴⁵ M. K. Such-Pyrgiel, D. Prokopowicz, A. Gołębiowska, *The postpandemic reality and the security of information technologies ICT, Big Data, Industry 4.0, social media portals and the Internet*, „Journal of Modern Science” 2022, vol. 49, no. 2, 21–22.

a) zwiększenie skali automatyzacji procesów, b) zwiększenie skali wydajności biznesu, c) zwiększenie możliwości przetwarzania dużych zbiorów danych i informacji, d) zwiększanie skali implementacji nowych modeli biznesowych opartych na dużej skali automatyzacji procesów wytwórczych itd. Jednak sztuczna inteligencja rozwijająca się w sposób niekontrolowany generuje poważne ryzyka, jak np.: zwiększenie skali dezinformacji, pojawiających się fake newsów, w tym banerów, memów zawierających spreparowane zdjęcia, grafiki, animacje, filmy prezentujące „fikcyjne fakty”, tj. bardzo realistycznie przedstawiające spreparowaną rzeczywistość. Pojawia się ryzyko potencjalnego zwiększenia skali wpływu na opinię publiczną, kształtowania ogólnospołecznej świadomości przez organizacje funkcjonujące bez poszanowania prawa. Z jednej strony niezbędne jest tworzenie zinformaty-zowanych i wystandaryzowanych narzędzi cyfrowych, diagnostycznych systemów informatycznych, zbudowanie wystandaryzowanego systemu oznaczeń informujących użytkowników, klientów, obywateli korzystających z określonych rozwiązań, produktów i usług, że są one wytworami sztucznej inteligencji, a nie człowieka. Z drugiej strony natomiast powinny być wprowadzone regulacje obligujące do informowania, że określona usługa lub produkt powstał jako wynik pracy wykonanej nie przez ludzi, ale przez sztuczną inteligencję. Wiele kwestii dotyczących właściwego społecznie, etycznie i biznesowo wykorzystywania technologii sztucznej inteligencji zostanie normatywnie uregulowanych w perspektywie kilku kolejnych lat⁴⁶.

14 czerwca 2023 r. Parlament Europejski uchwalił przełomowy w swojej istocie akt prawny regulujący stosowanie technologii sztucznej inteligencji. Jednak ponieważ owa technologia, w tym głównie sztuczna inteligencja generatywna, szybko się rozwija, a wprowadzenie sformułowanych w akcie przepisów jest planowane na lata 2026–2027, to z jednej strony podmioty gospodarcze wykorzystujące tę technologię mają sporo czasu, aby swoje procedury i produkty dostosować do wspomnianych regulacji⁴⁷, z drugiej

⁴⁶ A. Markiewicz, *Kogo dotyczy i co reguluje AI Act?*, <https://lexdigital.pl/kogo-dotyczy-i-co-reguluje-ai-act>, (27.06.2023).

⁴⁷ K. Krzywicka, *AI Act, czyli nowe wyzwania dla przedsiębiorców w erze rozwoju sztucznej inteligencji*, <https://www.parp.gov.pl/component/content/article/85510:ai-act-czyli-nowe-wyzwania-dla-przedsiębiorców-w-erze-rozwoju-sztucznej-inteligencji> (21.12.2023).

jednak nie można wykluczyć scenariusza, że mimo próby pełnego uregulowania rozwoju zastosowań tej technologii przez wdrożenie w 2027 r. ustawy dotyczącej właściwego, bezpiecznego i etycznego stosowania sztucznej inteligencji znów się okaże, że dynamiczny postęp technologiczny wyprzedza proces legislacyjny, który szybko rozwijających się technologii dotyczy⁴⁸.

Wobec powyższego w niniejszym rozdziale wskazano na zagrożenia wynikające z nieetycznego i nieuregulowanego prawnie zastosowania technologii sztucznej inteligencji. Wskazano także na znaczenie etyki odnoszącej się do rozwoju i zastosowań technologii sztucznej inteligencji w różnych dziedzinach działalności podmiotów gospodarczych i instytucji. Na podstawie opisanych przykładów nieetycznych zastosowań technologii AI wykazano, że z rozwojem technologii sztucznej inteligencji wiążą się zarówno aspekty pozytywne, jak i negatywne, tj. zarówno szanse rozwojowe, jak i zagrożenia dla rozwoju ludzkiej cywilizacji. Z jednej strony rozwój generatywnej sztucznej inteligencji jest swego rodzaju kolejnym etapem trwającego już od epoki pierwszej rewolucji przemysłowej postępu technologicznego, polegającego na wzroście skali uprzedmiotowienia pracy i automatyzacji procesów wytwórczych. Z drugiej strony rozwój ten generuje wiele nowych ryzyk i zagrożeń, których skalę można znacznie ograniczyć, jeżeli na każdym etapie procesu tworzenia, rozwoju i wdrażania technologii AI będą uwzględnione także kwestie etyczne, aspekty prospołeczne, a cały wspomniany proces zostanie uregulowany normatywnie, co powinno w dużym stopniu zmniejszyć skalę prawdopodobieństwa wymknięcia się tego rozwoju spod kontroli człowieka. W rozdziale wskazano na pierwsze potknięcia w próbie zastosowania sztucznej inteligencji na portalach internetowych mediów społecznościowych i opisano stosowane w ramach jej wykorzystania nieetyczne praktyki biznesowe oraz zagrożenia z tego wynikające. W rozdziale wskazano na potrzebę etycznego rozwoju i zastosowań technologii AI. Odbywający się w warunkach rywalizacji rynkowej czołowych korporacji technologicznych proces rozwoju technologii AI prawdopodobnie wkroczył już na ścieżkę bez odwrotu. Zapewne w kolejnych latach technologia AI będzie podlegała

⁴⁸ M. Szul, *AI Act – wszystko, co trzeba wiedzieć na temat ustawy o sztucznej inteligencji*, <https://www.ideoforce.pl/akademia/ai-act-wszystko-co-trzeba-wiedziec-na-temat-ustawy-o-sztucznej-inteligencji,841.html> (25.03.2024).

dalszemu rozwojowi i doskonaleniu. Będą się pojawiały nowe zastosowania coraz doskonalszej sztucznej inteligencji. Ważne jest więc, aby kontynuacja tego procesu postępu technologicznego nie wymknęła się spod kontroli człowieka. Aby tak się nie stało, aby nie zrealizowały się kreślone od pół wieku w powieściach i filmach *science fiction* czarne scenariusze katastrofy ludzkości spowodowanej niekontrolowanym rozwojem technologii AI, niezbędne jest uregulowanie prawne rozwoju tej technologii i nadanie wysokiego priorytetu aspektom etycznym i społecznym odnoszącym się do tego rozwoju.

Bibliografia

- 2001: *Odyseja kosmiczna – powieść*, [https://pl.wikipedia.org/wiki/2001:_Odyseja_kosmiczna_\(powiesc\)](https://pl.wikipedia.org/wiki/2001:_Odyseja_kosmiczna_(powiesc)) (15.01.2024).
- Amankwah-Amoah J., Khan Z., Wood G., Knight G., *COVID-19 and digitalization: The great acceleration*, „Journal of Business Research” 2021, no. 136, s. 602–611.
- Ashok M., Madan R., Joha A., Sivarajah U., *Ethical framework for artificial intelligence and digital technologies*, „International Journal of Information Management” 2022, no. 62, DOI: 10.1016/j.ijinfomgt.2021.102433.
- Baker Brunnbauer J., *Management perspective of ethics in artificial intelligence*, „AI and Ethics” 2021, no. 1(2), s. 173–181.
- Bellon M., *Stephen Hawking był geniuszem naszych czasów. Przed czym nas ostrzegal?*, 14 marca 2018, <https://businessinsider.com.pl/technologie/nauka/stephen-hawking-ostregal-swiat-przed-sztuczna-inteligencja/htl9zg2> (9.01.2024).
- Bieroński M., *Etyczne i moralne wyzwania związane ze stosowaniem Sztucznej Inteligencji*, „Kieleckie Studia Teologiczne” 2020, nr 19, s. 7–25.
- Braun J. von, Archer M. S., Reichberg G. M., Sánchez Sorond M., *AI, Robotics, and Humanity: Opportunities, Risks, and Implications for Ethics and Policy*, w: *Robotics, AI, and Humanity*, ed. J. von Braun et al., Springer, Cham 2021, s. 1–13.
- Cath C., *Governing Artificial Intelligence: Ethical, legal, and technical opportunities and challenges*, Royal Society, London, 2018.
- COVID-19 to plunge global economy into worst recession since World War*, World Bank, 2020, <http://www.worldbank.org/en/news/press-release/2020/06/08/covid-19-to-plunge-global-economy-intoworst-recession-since-world-war-ii> (15.07.2023).
- Czajkowski J., *Aitana – Wirtualna Influencerka Zmieniająca Reguły Gry w Marketingu Internetowym*, <https://idefi.pl/aitana-wirtualna-influencerka-zmieniajaca-reguly-gry-w-marketingu-internetowym> (5.12.2023).
- Díaz-Rodríguez N., Del Ser J., Coeckelbergh M., López de Prado M., Herrera-Viedma E., Herrera F., *Connecting the dots in trustworthy Artificial Intelligence: From AI*

- principles, ethics, and key requirements to responsible AI systems and regulation*, „Information Fusion” 2023, no. 99, doi.org/10.1016/j.inffus.2023.101896.
- Dmowski A., Sarnowski J., Prokopowicz D., *Podstawy finansów i bankowości*, Difin, Warszawa 2005.
- Goczał B., *Wirtualna Aitana zarabia 10 tys. euro miesięcznie. „Wirtualni influencerzy nie wyprą prawdziwych twórców”*, <https://www.wirtualnemedial.pl/artykul/wirtualna-influencerka-aitana> (3.12.2023).
- Grzegorek J., Prokopowicz D., Gwoździewicz S., *Zastosowanie zaawansowanych narzędzi przetwarzania danych w dobie cyfryzacji*, w: *Cyfryzacja w zarządzaniu*, red. A. Laskowska-Rutkowska, CeDeWu, Warszawa 2020, s. 93–128.
- Gupta M. sen, *What is digitization, digitalization, and digital transformation*, <https://www.arcweb.com/blog/what-digitization-digitalization-digital-transformation> (24.03.2020).
- Hermansyah M., Najib A., Farida A., Sacipto R., Rintyarna B. S., *Artificial Intelligence and Ethics: Building an Artificial Intelligence System that Ensures Privacy and Social Justice*, „International Journal of Science and Society” 2023, vol. 5, no. 1, s. 154–168.
- Huyue-Zhang A., *China generative AI regulation support growth*, <https://project-syndicate.org/commentary/china-generative-ai-regulation-support-growth-by-angela-huyue-zhang-2023-10> (10.2023).
- Interactive Charts on Artificial Intelligence*, <https://OurWorldInData.org/artificial-intelligence> (9.01.2024).
- Jakubczak W., Gołębiewska A., Prokopowicz D., *The Legal and Security Aspects of ICT and Industry 4.0 Importance for Financial Industry 4.0 Development*, „European Research Studies Journal” 2021, vol. 24, no. 4B, s. 169–181.
- Krzywicka K., *AI Act, czyli nowe wyzwania dla przedsiębiorców w erze rozwoju sztucznej inteligencji*, <https://www.parp.gov.pl/component/content/article/85510:ai-act-czyli-nowe-wyzwania-dla-przedsiębiorców-w-erze-rozwoju-sztucznej-inteligencji> (21.12.2023).
- Kwasek A., Maciaszczyk M., Kocot M., Rzepka A., Kocot D., Gąsiński H., Prokopowicz D., *Energy Saving Practices in the IT Area as a Factor of Sustainable Development of the Organization: A Case Study of Poland*, „Energies” 2023, no. 16(4) doi.org/10.3390/en16041942.
- Lem S., *Przyszłość otwarta na dobre i na złe*, „Znak” 2013, nr 700, za numerem: 1995, nr 484, s. 81–83, <http://www.miesiecznik.znak.com.pl/7002013stanislaw-lem-przyszlosc-otwarta-na-dobre-i-na-zle>.
- Lesandrini J., Idris M. Y., Reis D. S., *The Ethics of Artificial Intelligence and Machine Learning*, „Journal of Radiology Nursing” 2023, no. 42(3), s. 265–266.
- Łobejko S., *Digital transformation and innovativeness of enterprises*, „Optimum Economic Studies” 2020, no. 2 (100), DOI: 10.15290/oes.2020.02.100.03.
- Markiewicz A., *Kogo dotyczy i co reguluje AI Act?*, <https://lexdigital.pl/kogo-dotyczy-i-co-reguluje-ai-act> (27.06.2023).
- McCarthy J., Heyes P. J., *Some philosophical problems from the standpoint of artificial intelligence*, w: *Readings in artificial intelligence*, ed. B. L. Webber, N. J. Nilsson, Morgan Kaufmann 1969, s. 431–450.

- Owe A., Baum S. D., *Moral consideration of nonhumans in the ethics of artificial intelligence*, „AI and Ethics” 2021, no. 1(4), s. 517–528.
- Pizło W., Kulykovets O., Prokopowicz D., Mazurkiewicz-Pizło A., Kałowski A., Paprocka M. W., Stawicka E., Skarzyńska E., *The importance of Big Data Analytics technology in business management*, „Cybersecurity and Law” 2023, no. 2 (10), s. 270–282.
- Poselek M., *Zarabia 10 tysięcy euro, śledzi ją ponad 120 tysięcy osób. Problem w tym, że... nie istnieje*, 24.11.2023, <https://fashionbiznes.pl/aitana-lopez-kim-jest-wirtualna-influencerka>.
- Prokopowicz D., *Opportunities and threats to the development of Artificial Intelligence applications and the need for normative regulation of this development*, „International Journal of Legal Studies” 2023, no. 2 (14), s. 95–130.
- Prokopowicz D., *The postcovid rise in inflation: coincidence or the result of misguided, excessively interventionist and monetarist economic policies*, „International Journal of New Economics and Social Sciences” 2022, no. 16 (2) 2022, s. 105–148.
- Prokopowicz D., Gołębiowska A., Such-Pyrgiel M., *The role of Big Data and Data Science in the context of information security and cybersecurity*, „Journal of Modern Science” 2023, vol. 53, no. 4, s. 9–42.
- Ramaiah M., Chithanuru V., Padma A., Ravi V., *A Review of Security. Vulnerabilities in Industry 4.0 Application and the Possible Solutions Using Blockchain*, w: *Cyber Security Applications for Industry 4.0*, ed. R. Sujatha, G. Prakash, N. Zaman Jhanjhi, CRC Press, Boca Raton 2023, DOI: 10.1201/9781003203087-3.
- Rezolucja Parlamentu Europejskiego z dnia 16 lutego 2017 r. zawierająca zalecenia dla Komisji w sprawie przepisów prawa cywilnego dotyczących robotyki* (2015/2103 (INL)), Strasburg 2017.
- Roser M., *The brief history of artificial intelligence: The world has changed fast – what might be next?*, <https://ourworldindata.org/brief-history-of-ai> (6.12.2022).
- Różanowski K., *Sztuczna inteligencja: rozwój, szanse i zagrożenia*, „Zeszyty Naukowe Warszawskiej Wyższej Szkoły Informatyki” 2007, nr 2, s. 109–135.
- Savaram R., *Artificial Intelligence Vs Machine Learning*, <https://mindmajix.com/machine-learning-vs-ai> (17.12.2023).
- Sewpersadh N. S., *Disruptive business value models in the digital era*, „Journal of Innovation and Entrepreneurship” 2023, vol. 12, no. 2, doi.org/10.1186/s13731-022-00252-1.
- Stahl B. C., Brooks L., Hatzakis T., Santiago N., Wright D., *Exploring ethics and human rights in artificial intelligence – A Delphi study*, „Technological Forecasting and Social Change” 2023, no. 191, <https://doi.org/10.1016/j.techfore.2023.122502>.
- Stahl B. C., Wright D., *Ethics and privacy in AI and big data: Implementing responsible research and innovation*, „IEEE Security & Privacy” 2018, no. 16(3), s. 26–33.
- Stylec-Szromek P., *Sztuczna inteligencja – prawo, odpowiedzialność, etyka*, „Zeszyty Naukowe Politechniki Śląskiej” 2018, z. 123, s. 501–509.
- Such-Pyrgiel M. K., Prokopowicz D., Gołębiowska A., *The postpandemic reality and the security of information technologies ICT, Big Data, Industry 4.0, social media portals and the Internet*, „Journal of Modern Science” 2022, vol. 49, no. 2, s. 10–43.
- Szul M., *AI Act – wszystko, co trzeba wiedzieć na temat ustawy o sztucznej inteligencji*, <https://www.ideoforce.pl/akademia/ai-act-wszystko-co-trzeba-wiedziec-na-temat-ustawy-o-sztucznej-inteligencji,841.html> (25.03.2024).

- Talari S., *Why is Artificial Intelligence Important*, <https://mindmajix.com/why-ai-is-more-important-than-you-think> (17.08.2023).
- Turan T., Turan G., Kucuksille E. U., *Ethics of Artificial Intelligence: Impact on Society*, „The Journal of Graduate School of Natural and Applied Sciences of Mehmet Akif Ersoy University” 2022, no. 13(2), s. 292–299.
- Turski P., *Sztuczna inteligencja. „Musimy ją nauczyć myślenia w sposób ludzki”*, <https://trojka.polskieradio.pl/artykul/3003701,Sztuczna-inteligencja-Musimy-ja-nauczyc-myslania-w-sposob-ludzki> (19.07.2022).
- Tzimas T., *Legal and Ethical Challenges Artificial Intelligence from an International Law Perspective*, Springer Nature Switzerland AG, Switzerland 2021.
- Warner J., *Depression or inflation: where will Covid-19 end?*, 2020, <http://www.telegraph.co.uk/business/2020/04/12/depression-inflation-will-covid-19-end> (17.07.2023).
- Yadav N., *Ethics of Artificial Intelligence and Robotics: Key Issues and Modern Ways to Solve Them*, „Journal of Digital Technologies and Law” 2023, no. 1(4), s. 955–972.
- Yerukala M., *Artificial Intelligence Tutorial*, <https://mindmajix.com/artificial-intelligence-tutorial> (16.12.2023).

Risks of unethical and unregulated use of artificial intelligence technology

Summary

The main purpose of the chapter was to point out the dangers of unethical and unregulated applications of artificial intelligence technologies. The importance of ethics relating to the development and application of artificial intelligence (AI) technology in various fields of business entities and institutions was also pointed out. There are both positive and negative aspects associated with the development of artificial intelligence technology, i.e. both development opportunities and threats to the development of human civilization. On the one hand, the development of generative artificial intelligence is a kind of next stage of the ongoing technological progress, which has been taking place since the era of the first industrial revolution, consisting in the increase in the scale of objectification of labor and automation of manufacturing processes. On the other hand, this development is associated with a number of risks and dangers, the scale of which can be significantly reduced if at each stage of the process of creation, development and implementation of AI technologies, ethical issues, pro-social aspects are also taken into account and the whole said process is normatively regulated, which should greatly reduce the scale of the likelihood of this development slipping out of human control. The chapter points out the first stumbling blocks in the attempt to apply artificial intelligence to online social media portals, and describes the unethical business practices used in the use of artificial intelligence technology on online social media portals and

the risks arising from this. The chapter points out the need for ethical development and applications of AI technologies.

Keywords: artificial intelligence, generative artificial intelligence, applications of artificial intelligence, ethics of artificial intelligence, legal norms of artificial intelligence, dangers of artificial intelligence development, Internet, social media

Copyright by Wydawnictwo Naukowe UKSW

Rozwój i zastosowania technologii sztucznej inteligencji z uwzględnieniem aspektów etycznych i społecznych

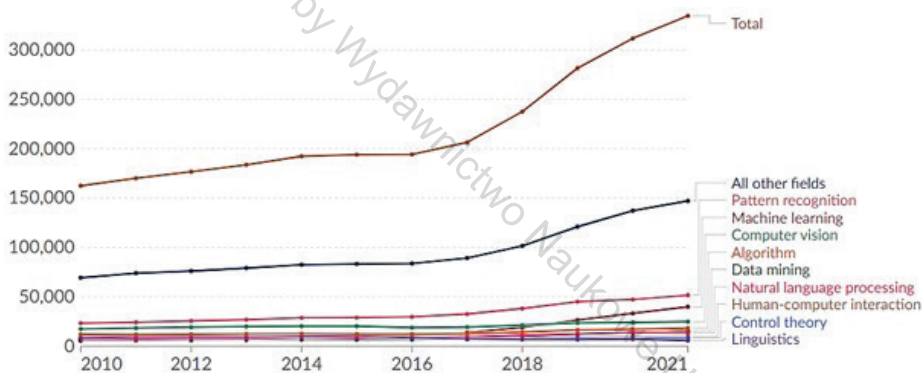
Wprowadzenie

Według założyciela Microsoftu Billa Gatesa dokonujący się obecnie rozwój technologii algorytmów generatywnych, tj. kluczowej dziedziny sztucznej inteligencji, to kolejny duży skok technologiczny w kontekście postępu technologicznego ostatnich kilku dekad. Gates twierdzi, że chatboty oparte na algorytmach generatywnych stanowią rozwiązanie innowacyjne, porównywalne z tym, jakim w przeszłości okazały się mikroprocesor, komputer osobisty, internet, telefon komórkowy czy graficzny interfejs użytkownika (GUI). Opracowanie GUI było w ramach rozwoju komputerów osobistych rozwiązaniem rewolucyjnym, ponieważ pozwoliło na stworzenie w latach 80. XX wieku systemów operacyjnych Windows i Mac OS, umożliwiającym korzystanie z komputerów użytkownikom nieposiadającym zaawansowanej wiedzy informatycznej. Dzięki wyposażeniu komputerów osobistych w systemy operacyjne oparte na GUI każdy użytkownik mógł szybko intuicyjnie opanować obsługę komputera, ponieważ wpisywane wcześniej kody i polecenia zastąpiono obrazkami, okienkami i ikonami, które swoim wyglądem

¹ Dr Dariusz Prokopowicz, Instytut Ekonomii i Finansów, Wydział Społeczno-Ekonomiczny UKSW.

symbolizowały określone funkcje obsługi systemu operacyjnego. W związku z powyższym już w nieodległej przyszłości sztuczna inteligencja zmieni zdaniem Billa Gatesa sposób, w jaki ludzie pracują, uczą się, podróżują, korzystają z opieki zdrowotnej i komunikują się ze sobą². Dokonujący się w ostatnich latach wzrost skali zastosowań technologii AI w różnych dziedzinach działalności firm, przedsiębiorstw i instytucji skorelowany jest także z dużym wzrostem liczby publikacji naukowych, w których opisano wyniki badań naukowych i różne aspekty rozwoju technologii sztucznej inteligencji oraz jej zastosowań. Na poniższym wykresie przedstawiono publikowane w skali poszczególnych lat publikacje naukowe na temat sztucznej inteligencji według dziedzin studiów.

Wykres 1. Coroczne publikacje naukowe na temat sztucznej inteligencji według dziedzin studiów (w skali globalnej)



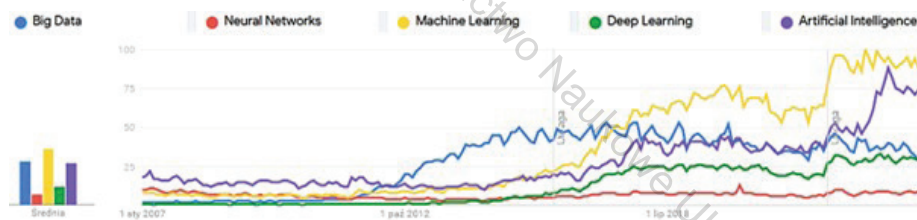
Publikacje naukowe w języku angielskim związane z rozwojem i zastosowaniem sztucznej inteligencji. Wykres ten obejmuje artykuły z czasopism, artykuły konferencyjne, publikacje z repozytoriów (takie jak arXiv), książki i rozprawy. Zgłoszona liczba publikacji w 2021 r. może być niższa niż liczba rzeczywista ze względu na opóźnienia w gromadzeniu danych.

Źródło: *Annual scholarly publications on AI, by field of study*, <https://ourworldindata.org/grapher/number-ai-publications-by-field> (20.12.2023)

² *Bill Gates o rewolucyjnej technologii. „Zmieni sposób, w jaki ludzie pracują, uczą się, podróżują”*, 22.03.2023, <https://tvn24.pl/biznes/ze-swiata/chat-gpt-bill-gates-o-rewolucyjnej-technologiei-zmieni-sposob-w-jaki-ludzie-pracuja-ucza-sie-podrozuja-6854903>.

Na podstawie internetowej analityki przeprowadzonej z wykorzystaniem googlowego narzędzia Google Trends potwierdzono wysoki poziom korelacji wyszukiwanych w internecie haseł dotyczących wybranych kluczowych zagadnień związanych z problematyką sztucznej inteligencji w okresie ostatnich kilkunastu lat. Wysoki poziom korelacji wystąpił między takimi hasłami jak *artificial intelligence*, *machine learning*, *deep learning* i w nieco mniejszym stopniu *neural networks*. Poniższy wykres prezentuje wynik analityki Google Trends w zakresie zmiany poziomu zainteresowania internautów problematyką sztucznej inteligencji, przeprowadzonej na podstawie wybranych haseł wpisywanych do wyszukiwarki Google w okresie od 1 stycznia 2007 do 18 stycznia 2024 r. w skali globalnej. Wykres przedstawia wzrost zainteresowania internautów odnośnie problematyką od 2015 r. Jedynie wzrost zainteresowania zagadnieniami związanymi z hasłem *big data*³ rozpoczął się ok. 3 lat wcześniej względem kluczowych zagadnień sztucznej inteligencji.

Wykres 2. Analiza zainteresowania internautów wybranymi zagadnieniami związanymi z problematyką sztucznej inteligencji, przeprowadzona na podstawie wybranych haseł wpisywanych do wyszukiwarki Google w okresie od 1 stycznia 2007 do 18 stycznia 2024 r. (globalnie)



Źródło: Opracowanie własne na podstawie analityki Google Trends (17.08.2023)

Podstawowe zalety rozwoju sztucznej inteligencji to przede wszystkim wzrost efektywności procesów wielokryterialnego, złożonego, zaawansowanego przetwarzania dużych zbiorów danych i wynikający z tego wzrost zastosowań AI w różnych sektorach i branżach gospodarki, w różnych procesach

³ S. Gwoździewicz, D. Prokopowicz, *Determinants of electronic data interchange security in the context of Big Data and cloud computing technology development. Sentiment analysis conducted for cybercrime issues occurred in the period from May 2017 to February 2019*, „International Journal of New Economics and Social Sciences” 2022, vol. 15, no. 1, s. 93–94.

wytwórczych, w zarządzaniu złożonymi procesami, w oferowanych usługach itd. Do negatywnych aspektów owego rozwoju zalicza się natomiast rosnące ryzyko wymknięcia się go spod kontroli człowieka. Z negatywnymi aspektami związane są także obawy ludzi w kwestii zmian na rynkach pracy w przyszłości. Zmiany te będą polegać na zastępowaniu pracy ludzkiej przez roboty wyposażone w sztuczną inteligencję. Najpierw tego typu zmiany dokonają się w zakresie czynności powtarzalnych, niewymagających kreatywności, innowacyjności, artyzmu itp. Jednak już obecnie powstają roboty wyposażone w sztuczną inteligencję, technologie uczącej się maszyny, a także w algorytmy umożliwiające wykonywanie przez nie prac wymagających zastosowania także kreatywnych rozwiązań i/lub innowacyjnych pomysłów. Główne dziedziny zastosowań sztucznej inteligencji przedstawia poniższy rysunek.

Rysunek 1. Główne dziedziny zastosowań sztucznej inteligencji



Źródło: M. Yerukala, *Artificial Intelligence Tutorial*, <https://mindmajix.com/artificial-intelligence-tutorial> (16.12.2023)

Do dynamicznie rozwijających się obecnie zastosowań sztucznej inteligencji realizowanych w czasie rzeczywistym zaliczają się: chatboty, wirtualni asystenci osobiści, stosowane w handlu internetowym systemy rekomendacji, wykrywanie oszustw, pojazdy autonomiczne, przetwarzanie języka naturalnego (NLP), rozpoznawanie obrazów i obiektów, diagnoza

opieki zdrowotnej, systemy inteligentnych domów oraz handel finansowy⁴. W kolejnych latach, wraz z rozwojem i sukcesywnym doskonaleniem technologii sztucznej inteligencji, w tym sztucznej inteligencji generatywnej, możliwości wykorzystywania jej do automatyzacji procesów i do prac wymagających wysokiego poziomu kreatywności, innowacyjności i artyzmu będą się zwiększać. Już obecnie generatywna sztuczna inteligencja tworzy grafiki, zdjęcia, plakaty, obrazy i animacje zawierające elementy artyzmu, czyli coś, co do niedawna było domeną działalności wyłącznie ludzkiej. Z jednej strony pojawiają się nowe możliwości wykorzystania generatywnej AI jako nowego narzędzia usprawniającego pracę grafików komputerowych i filmowców, z drugiej – także kontrowersje dotyczące aspektów etycznych⁵ i niezbędnych uregulowań prawnych w zakresie praw autorskich do dzieł sztuki tworzonych w ten sposób.

Kontrowersyjne orzeczenie sądowe sądu internetowego w Pekinie a aspekty etyczne

Uregulowanie prawne kwestii zastosowań generatywnej sztucznej inteligencji i tym samym także jej rozwoju odbywa się na poziomie niektórych krajów i Unii Europejskiej. Niestety nie ma w tej kwestii unifikacji globalnej, mimo tego że rozwój tego rodzaju technologii i jej zastosowań, np. w zakresie doskonalenia określonych aplikacji internetowych, ma charakter coraz bardziej globalny. W niektórych krajach, jak np. w Chinach, podejście do tematu uregulowania prawnego zastosowań technologii sztucznej inteligencji jest znacznie bardziej liberalne względem np. praktykowanego w Unii Europejskiej. Zróżnicowanie w tej kwestii może generować także inne rozstrzygnięcia sądowe, dotyczące np. uznania lub nieuznania praw autorskich przyznanych osobom wykorzystującym aplikacje internetowe lub systemy informatyczne

⁴ M. Yerukala, *Artificial Intelligence (AI) Projects*, <https://mindmajix.com/artificial-intelligence-projects> (17.08.2023).

⁵ C. Cath, *Governing Artificial Intelligence: Ethical, legal, and technical opportunities and challenges*, Royal Society, London 2018, s. 39.

wyposażone w określone rozwiązania generatywnej sztucznej inteligencji, które występują w roli twórcy tworzącego swego rodzaju dzieła kultury i/lub dzieła sztuki w formie grafik, zdjęć, animacji, filmów, opowiadań, wierszy itd., mających cechy unikalności i niepowtarzalności. Jednym z przykładów kontrowersyjnego orzeczenia sądowego dotyczącego wspomnianej kwestii było orzeczenie wydane przez jeden z sądów w Chinach.

Otóż w październiku 2023 r. sąd internetowy w Pekinie wydał swego rodzaju kontrowersyjne i jednocześnie przełomowe orzeczenie, w którym potwierdzone zostały prawa autorskie do dzieł sztuki tworzonych z zastosowaniem sztucznej inteligencji. Wspomniane orzeczenie jest zasadniczo odmienne względem wydanego nieco wcześniej orzeczenia Urzędu Patentów i Znaków Towarowych Stanów Zjednoczonych (United States Patent and Trademark Office), która to instytucja nie uznała praw autorskich do takich dzieł. W świetle wspomnianego orzeczenia sądu pekińskiego zwolennicy liberalnego podejścia do wszelkiego rodzaju uregulowania kwestii zastosowań nowych technologii mogą uznać, że Chiny stosują wobec technologii sztucznej inteligencji strategię prowzrostową, przyjazną dla zastosowań biznesowych, a polegającą na tym, aby kwestia zastosowań technologii AI została uregulowana w stopniu możliwie znikomym, tak by zminimalizować ograniczenia rozwoju tej technologii⁶.

Istotnym zagadnieniem jest także to, jakie przesłanki doprowadziły do wydania wspomnianego orzeczenia przez sąd internetowy w Pekinie. Otóż sąd ten wydał w pierwszej instancji wyrok w sprawie naruszenia praw autorskich dotyczących obrazów generowanych z zastosowaniem technologii AI. Sporna kwestia trafiła do sądu, ponieważ powód, pan Li, zarzucił blogerowi Bai Jia Hao użycie w artykule grafiki, którą uprzednio stworzył pan Li, i opublikowanie tegoż artykułu bez zgody pana Li na internetowej platformie Xiaohongshu. Do stworzenia przedmiotowej grafiki pan Li zastosował określony system generatywnej sztucznej inteligencji. W ten sposób zaistniał precedens dotyczący naruszenia praw autorskich do obrazów, grafik lub innych dzieł generowanych z zastosowaniem informatycznych narzędzi wyposażonych w technologię AI lub inaczej mówiąc: dzieł, które mają cechy unikalności,

⁶ A. Huyue-Zhang, *China generative AI regulation support growth*, <https://project-syndicate.org/commentary/china-generative-ai-regulation-support-growth-by-angela-huyue-zhang-2023-10>.

mogą być uznane za wyjątkowe w swym rodzaju, mogą być uznane za artystyczne i zostały stworzone przez generatywną sztuczną inteligencję.

W związku z powyższym co w swej istocie orzekł sąd? Otóż sąd uznał, że dzieło w formie grafiki stworzonej przez informatyczną aplikację wyposażoną w generatywną sztuczną inteligencję ma cechy oryginalności, unikalności, niepowtarzalności itp., jest elementem twórczej działalności człowieka posługującego się technologią AI i wytworzone w ten sposób dzieło odzwierciedla intelektualny wkład człowieka. Niezależnie od tego, w jakim zakresie człowiek posłużył się w swoim akcie tworzenia technologią AI, powstałe w ten sposób dzieło, będące unikalnym, niepowtarzalnym i w pewnym sensie także „oryginalnym”, powinno zostać uznane za podlegające ochronie prawa autorskiego. A więc wspomniany sąd potraktował generatywną sztuczną inteligencję jako kolejne z wielu narzędzi, jakie w ramach dokonującego się postępu technologicznego stworzył człowiek. Sąd uznał technologię AI jako jedno z wielu narzędzi, które człowiek wykorzystuje w roli twórcy nowych dzieł, a nie jako osobny kreatywny byt. Sąd uznał, że od momentu zaprojektowania dzieła przez pana Li do momentu wyboru ostatecznej formy dzieła człowiek – twórca miał znaczący wkład intelektualny. Ów wkład, który miał duży wpływ na ostateczną formę dzieła – grafiki, stanowiło m.in. zaprojektowanie sposobu prezentacji postaci, wybór słów-kluczy, ustalenie ich kolejności, ustawienie parametrów, wybór odpowiedniego obrazu itd. A więc powstałe w ten sposób dzieło jest odzwierciedleniem intelektualnego wkładu pana Li i w związku z tym to człowiek jest autorem grafiki i posiada do tego dzieła prawa autorskie.

Teoria równowagi w zakresie przyznawania decyzyjności inteligentnym systemom informatycznym

Zastosowanie przez pana Li technologii AI podczas tworzenia grafiki, której wykorzystanie i zamieszczenie na stronie internetowej przez blogera Bai Jia Hao może nie powodować uszczerbku na zdrowiu lub innych strat dla osób trzecich, to jedna kwestia. Inna to np. stworzenie nowego leku na

określone schorzenia, w tym choroby śmiertelne⁷, na bazie opracowanej przez technologię AI unikalnej receptury różnych związków chemicznych. Taki medykament może powodować wcześniej nieznanie bardzo negatywne skutki uboczne⁸. Kto będzie odpowiedzialny za ewentualne błędy proceduralne powstałe w procesie tworzenia leku, kogo obciąży kosztami firma ubezpieczeniowa? Instytut badawczy, który zlecił opracowanie leku podwykonawcy, którym jest firma będąca właścicielem inteligentnego systemu informatycznego wyspecjalizowanego w tworzeniu unikalnych receptur leków, czy może raczej firmę informatyczną, która stworzyła określony system generatywnej sztucznej inteligencji zaimplementowany do inteligentnego systemu informatycznego opracowującego receptury leków? Czy etyczne jest zrzucanie odpowiedzialności za ewentualne szkody na narzędzia technologiczne, które wcześniej stworzył przecież także człowiek? Czy etyczne jest angażowanie technologii AI i zlecenie inteligentnemu systemowi informatycznemu opracowania receptury nowego leku w sytuacji, gdy nie może być w pełni rozstrzygnięta kwestia ani autorstwa tego rodzaju „dzieła”, jakim jest opracowanie unikalnej receptury nowego leku, ani odpowiedzialności za ewentualne popełnione błędy, szkody itp.⁹

Z rozstrzygnięciem tego rodzaju kwestii wiązą się zakres odpowiedzialności za realizowane zleczone zadania oraz zakres decyzyjności, jaki jest obecnie i jaki zostanie przyznany w przyszłości systemom informatycznym, maszynom i autonomicznym robotom wyposażonym w sztuczną inteligencję. Jednak zakres tej decyzyjności może zależeć od wielu czynników, jak np. czasu, w jakim decyzja powinna być podjęta. Na przykład inteligentne systemy informatyczne w bankach komercyjnych przeprowadzają badanie zdolności kredytowej potencjalnych kredytobiorców składających wnioski kredytowe. Sztuczna inteligencja może w oparciu o wiele danych

⁷ A. Olewiński, *AI w białym fartuchu. Sztuczna inteligencja i nowe leki. Jak jest i jak będzie wykorzystywana w Polsce, Europie i na świecie?*, 24.07.2023, <https://aptekarski.com/arttykul/ai-w-bialym-fartuchu-sztuczna-inteligencja-i-nowe-leki-jak-jest-i-jak-bedzie-wykorzystywana-w-polsce-europie-i-na-swiecie>.

⁸ A. Kuchta, *Sztuczna inteligencja przyspiesza powstanie leków*, 16.10.2023, <https://ai.infor.pl/sztuczna-inteligencja/6338561,sztuczna-inteligencja-przyspiesza-powstanie-lekow.html>.

⁹ K. Murphy, E. Di Ruggiero, R. Upshur, D. J. Willison, N. Malhotra, J. C. Cai, J. Gibson, *Artificial intelligence for good health: a scoping review of the ethics literature*, „BMC Medical Ethics” 2021, no. 22(1), s. 5–6.

dotyczących osoby wnioskującej o kredyt: jej stanu majątkowego, charakterystyki i poziomu dochodów, wartości potencjalnych zabezpieczeń kredytu itd., precyzyjnie i szybko oszacować zdolność do spłaty kredytu w określonym czasie z uwzględnieniem konkretnego oprocentowania i innych kosztów związanych z korzystaniem z tego rodzaju zewnętrznego finansowania. Jednak biorąc pod uwagę obowiązujące regulacje prawne i standardy procedur kredytowych, jest mało prawdopodobne, aby w nieodległej perspektywie czasu instytucja nadzoru bankowego, tj. Komisja Nadzoru Finansowego, dopuściła do możliwości podejmowania przez bankowe systemy informatyczne wyposażone w technologię AI decyzji o udzieleniu lub nieudzieleniu kredytu bankowego. Systemy bankowe mogą przeprowadzić analizę zdolności kredytowej i wydać rekomendację dla analityka kredytowego lub komitetu kredytowego¹⁰. Czas podejmowania decyzji jest dla banku i dla klienta istotny, ale jest to czas mierzony w dniach lub godzinach, a nie w sekundach.

Gdyby wyposażony w technologię AI system informatyczny sterujący procesem prowadzenia autonomicznego samochodu¹¹, poruszającego się po drogach w trybie „bez kierowcy”, miał podejmować decyzje w sytuacjach wysokiego ryzyka kolizji z innymi pojazdami, obiektami czy ludźmi znajdującymi się na drodze lub w jej pobliżu, czas reakcji, i tym samym podejmowania decyzji, byłby liczony w sekundach lub ułamkach sekund. Jeśliby ta decyzja miała polegać na wyborze akceptacji kolizji z innym pojazdem nadjeżdżającym z naprzeciwka celem ominięcia ludzi, którzy przypadkowo znaleźli się na jezdni, system musiałby tę decyzję podjąć jednak sam, ponieważ czasu na poinformowanie o potencjalnych ryzykach wynikających z nietypowej sytuacji, jaka wystąpiła na drodze, i przekazanie decyzyjności ludziom znajdującym się w autonomicznym samochodzie raczej by już nie było¹². W związku z tym producenci pojazdów autonomicznych nie zdecydowali się na przekazanie w pełnym zakresie wspomnianej decyzyjności

¹⁰ A. Dmowski, J. Sarnowski, D. Prokopowicz, *Podstawy finansów i bankowości*, Difin, Warszawa 2005, s. 263.

¹¹ *Robot ethics 2.0: From autonomous cars to artificial intelligence*, ed. P. Lin, K. Abney, R. Jenkins, Oxford University Press, Oxford 2017, s. 37.

¹² P. Stylec-Szromek, *Sztuczna inteligencja...*, art. cyt., s. 508.

inteligentnym systemom informatycznym odpowiedzialnym za prowadzenie tego rodzaju pojazdów po drogach, ulicach, autostradach itp. Nadal człowiek znajdujący się wewnątrz autonomicznie poruszającego się pojazdu jest zobligowany do kontroli jazdy sterowanej przez system wyposażony w technologię AI.

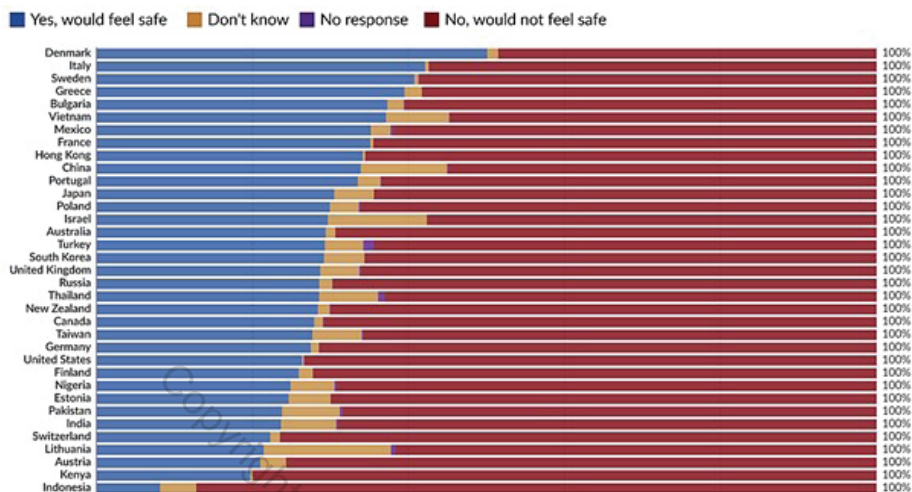
Wobec powyższego z przedstawionych powyżej dwóch przykładów różnych zastosowań technologii AI oraz z przyznania systemom informatycznym wyposażonym w te technologie określonego zakresu decyzyjności, z uwzględnieniem kwestii podejmowania decyzji w warunkach nietypowych sytuacji generujących różne ryzyka, jednoznacznie wynika, że zakres decyzyjności przyznanej inteligentnym aplikacjom, systemom informatycznym, robotom itp. zależy od wielu determinantów i specyfiki możliwych zdarzeń oraz nietypowych sytuacji, w jakich decyzje są podejmowane. Zapewne znaczna część tych determinantów będzie wynikać z aspektów technicznych i statystycznych, a nie tylko normatywnych i etycznych¹³.

Poziom ogólnospołecznej świadomości obywateli dotyczącej kwestii bezpieczeństwa jazdy samochodem autonomicznym jest w poszczególnych krajach zróżnicowany. Zróżnicowanie to przedstawiono na kolejnym wykresie na przykładzie wybranych krajów.

Z wykresu tego wynika, że w wielu krajach większość obywateli jest przekonana, że z jazdą samochodem autonomicznym wiązą się liczne zagrożenia, tj. że jazda pojazdem autonomicznym w trybie „bez kierowcy” nie jest bezpieczna. Poziom świadomości społecznej w tej kwestii koreluje więc ze statystykami wciąż wysokiego poziomu awaryjności inteligentnych systemów kontroli jazdy zastosowanych w tego rodzaju pojazdach i nadal z tych przyczyn się zdarzających poważnych wypadków drogowych.

¹³ H. Yusep Multana, *Ethics and Law for the Use and Development of Artificial Intelligence Technology*, „East Asian Journal of Multidisciplinary Research” 2023, vol. 2, no. 11, s. 4523–4524.

Wykres 3. Poglądy na temat bezpieczeństwa jazdy samochodem autonomicznym (badanie z 2021 r., przeprowadzone w wybranych krajach)



Respondenci ankiety zostali zapytani: „Czy czułbyś się bezpiecznie w samochodzie sterowanym przez komputer, bez kierowcy?”

Źródło: Interactive Charts on Artificial Intelligence, <https://OurWorldInData.org/artificial-intelligence> (9.01.2024)

W świetle zaprezentowanych wyżej rozważań zaproponowana przez autora niniejszego rozdziału teoria równowagi w zakresie przyznawania decyzyjności inteligentnym systemom informatycznym, maszynom i autonomicznym robotom wyposażonym w sztuczną inteligencję polega na odpowiednio symetrycznym skorelowaniu wspomnianej decyzyjności z przeniesionym na smarttechnologie i systemy informatyczne zakresem odpowiedzialności za realizowane zlecone zadania oraz na uwzględnieniu etycznego stosowania technologii AI i kontroli człowieka nad poziomem rzetelności etycznego rozwoju zastosowań AI. W związku z tym zapewnienie etycznego stosowania technologii AI powinno być zdeterminowane efektywnie działającym systemem, zapewniającym człowiekowi kontrolę nad działaniem inteligentnych systemów informatycznych robotów wyposażonych w sztuczną inteligencję, którym powierzono sterowanie precyzyjnie zdefiniowanymi procesami, wykonywanie zadań według możliwych wariantów scenariuszy sytuacyjnych i zaprogramowanych algorytmów możliwych zdarzeń oraz przydzielono określony zakres decyzyjności. Wspomniana kontrola powinna uwzględniać

także kwestie etyczne i społeczne zgodnego z potrzebami ludzi wykorzystania technologii AI w jej różnych zastosowaniach. Tylko w ten sposób zastosowania technologii sztucznej inteligencji mogą przynieść ludzkości więcej korzyści wobec generowanych aspektów negatywnych. Jeżeli tak się stanie, to technologia generatywnej sztucznej inteligencji powinna w przyszłości pomóc w rozwiązywaniu kluczowych problemów rozwoju cywilizacji, w tym np. w zakresie opracowania nowych terapii medycznych, skuteczniejszego leczenia chorób onkologicznych, stworzeniu nowych odmian roślin uprawnych, bardziej odpornych na różne negatywne czynniki środowiska, opracowaniu nowych, bardziej wydajnych zielonych technologii wytwarzania energii z odnawialnych i bezemisyjnych źródeł energii itd., a także problemów wynikających z nadal generowanego przez człowieka wysokiego poziomu zanieczyszczenia środowiska i emisji gazów cieplarnianych. Już obecnie technologie Przemysł 4.0/5.0¹⁴, w tym Big Data Analytics¹⁵ i generatywnej sztucznej inteligencji, są wykorzystywane do doskonalenia z informatyzowanych, analitycznych modeli prognostycznych długoterminowych zmian klimatu¹⁶.

Etyka a kwestia praw autorskich do dzieł tworzonych z zastosowaniem generatywnej sztucznej inteligencji

Kontynuując rozważania z poprzedniego podrozdziału, należy wspomnieć, że kwestie decyzyjności przedstawiają się nieco inaczej, gdy dotyczą np. rozstrzygnięcia o autorstwie danego dzieła, uznania praw autorskich do

¹⁴ A. Gołębiowska, D. Prokopowicz, *Growing importance of ICT, Industry 4.0 and Big Data Analytics as key determinants of the development of The Financial Industry 4.0*, w: *Socio-economic and legal dimensions of digital transformation. Selected contexts*, ed. A. Gołębiowska, M. Such-Pyrgiel, Wydawnictwo SGSP, Warszawa 2021, s. 35–56.

¹⁵ W. Pizło, O. Kulykovets, D. Prokopowicz, A. Mazurkiewicz-Pizło, A. Kałowski, M. W. Paprocka, E. Stawicka, E. Skarzyńska, *The importance of Big Data Analytics technology in business management*, „Cybersecurity and Law” 2023, no. 2 (10), s. 277–278.

¹⁶ B. Day, *Basic Concepts of Big Data and Artificial Intelligence in Managing Climate Data*, <https://www.climate4life.info/2018/12/kompasi-dasar-big-data-dan-artificial-intelligence-dalam-pengelolaan-data-iklim.html> (4.11.2023).

obrazu, grafiki, filmu, tekstu, innowacyjnego rozwiązania, receptury leku, patentu itd., powstałego jako niepowtarzalne, oryginalne dzieło w ramach aktu tworzenia, do którego twórca, człowiek, zaangażował technologię AI, odgrywając rolę wykonawcy zlecenia stworzenia czegoś nowego. Tego rodzaju rozstrzygnięcia powinny się opierać przede wszystkim na unormowaniach prawnych precyzyjnie definiujących to, czym w swojej istocie jest akt tworzenia czegoś nowego z zastosowaniem sztucznej inteligencji. Jednak w celu zdiagnozowania kwestii oryginalności powstałych w ten sposób dzieł może być niezbędna także technologia, która będzie określać poziom innowacyjności, artyzmu i indywidualności efektu „pracy” wykonanej przez inteligentną aplikację informatyczną przysposobioną w procesie głębokiego uczenia¹⁷ do tworzenia określonego rodzaju dzieł.

Rozważania dotyczące rozstrzygnięcia kwestii praw autorskich do dzieł tworzonych z zastosowaniem generatywnej sztucznej inteligencji dalej się toczą w dyskusjach prowadzonych w środowiskach naukowych, co jest zdeterminowane różnymi czynnikami, do których zalicza się m.in. kwestie poziomu artyzmu, istoty pojęcia artyzmu, relacji człowiek–maszyna, projekt dzieła–jego realizacja z zastosowaniem określonych narzędzi i technologii itd. Jeżeli np. inwestor zleca wykonanie dzieła mającego cechy artyzmu, wskazuje, co konkretnie ma powstać, to kto jest twórcą, skoro wykonawca tworzy unikalne, niepowtarzalne opracowanie określonego projektu dzieła lub dzieło zostało stworzone według niepowtarzalnego, oryginalnego projektu? Czy w takich sytuacjach twórcą zawsze będzie raczej wykonawca zlecenia o określonych cechach artyzmu czy będzie nim raczej zleceniodawca i jednocześnie inwestor? Jeżeli zleceniodawca zleci malarzowi namalowanie obrazu lub fotografowi wykonanie zdjęć według określonego, niepowtarzalnego projektu, a dzieło będzie posiadało także swego rodzaju cechy artystycznego, oryginalnego wykonania, to kto będzie wskazany jako twórca i właściciel praw autorskich do stworzonego dzieła lub określonych jego części, cech czy atrybutów?

W tej kwestii zapewne nie zawsze będzie funkcjonować jedna stała zasada wynikająca z zapisów umowy regulującej m.in. zdefiniowane i przyznane

¹⁷ J. Lesandrini, M. Y. Idris, D. S. Reis, *The Ethics of Artificial Intelligence and Machine Learning*, „Journal of Radiology Nursing” 2023, no. 42(3), s. 265–266.

prawa autorskie zleceniodawcy lub wykonawcy dzieła mającego cechy sztuki. W związku z tym, gdy w trakcie realizacji określonego dzieła zostanie zaangażowana technologia generatywnej sztucznej inteligencji, to kwestią kluczową będzie to, kto jest właścicielem tej technologii, kto był twórcą narzędzia funkcjonującego jako inteligentny system informatyczny wykonujący określone dzieła zgodnie z wcześniej zaprogramowanymi algorytmami, zgodnie z wyuczonymi procesami realizacji określonych czynności, zgodnie z precyzyjnie sformułowanymi wytycznymi dotyczącymi efektu finalnego dzieła, które ma powstać itd. Wobec powyższego to, jak będą realizowane aspekty etyczne¹⁸ w kontekście rozstrzygnięcia kwestii sztuki dzieł tworzonych z zastosowaniem generatywnej sztucznej inteligencji oraz kwestii praw autorskich do takich dzieł, będzie zależało także od tego, czy istota sztuki będzie wiązana z aktem tworzenia realizowanym wyłącznie przez człowieka oraz czy technologia generatywnej sztucznej inteligencji tworzącej unikalne, niepowtarzalne i oryginalne dzieła artystyczne zostanie uznana za swego rodzaju „artystę” czy jedynie za narzędzie w rękach artysty, którym może być tylko człowiek. Kwestia ta sprowadza się więc do istoty sztuki.

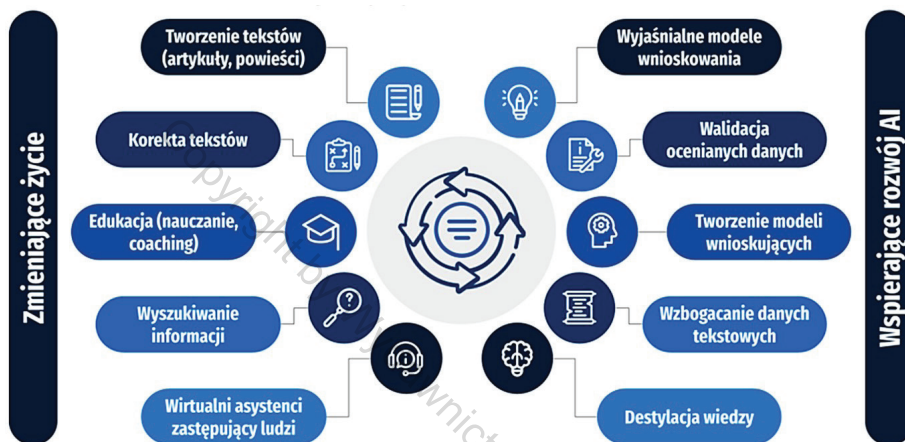
W najbliższych latach kwestia uznania praw autorskich do dzieł tworzonych z zastosowaniem technologii generatywnej sztucznej inteligencji może jeszcze w poszczególnych krajach z różnych względów wyglądać niejednakowo. Zastosowanie rozlicznych inteligentnych systemów informatycznych, różny poziom unikalności i sztuki, wzorowanie się lub czerpanie inspiracji z innych dzieł, będących m.in. wytworami ludzkiej aktywności, znajdujących rozmaite zastosowania, to tylko niektóre determinanty różniące. Kwestią szczególnie istotną w kontekście praw autorskich są także zakres i jakość wykazywania źródeł, tj. materiałów, tekstów, grafik, zdjęć lub innych, które zostały wytworzone z zastosowaniem genAI. Dostępne obecnie w internecie inteligentne chatboty, typu stworzony przez firmę OpenAI ChatGPT, Microsoftowy Copilot, Grok firmy (i jednocześnie internetowego medium społecznościowego) X (dawny Twitter), nie wykazują niestety w pełnym zakresie wszystkich źródeł inspiracji, cytowań, odwołań itd., czyli nie podają pełnych opisów materiałów i publikacji źródłowych

¹⁸ J. C. Bélisle Pipon, E. Monteferrante, M. C. Roy, V. Couture, *Artificial intelligence ethics has a black box problem*, „AI & Society” 2022, s. 7–8.

jako informacji uzupełniającej do nowo tworzonych dzieł z zastosowaniem inteligentnych chatbotów lub innych z informatyzowanych narzędzi wyposażonych w sztuczną inteligencję.

Kluczowe obszary perspektywiczne zastosowań ChatGPT przedstawiono na poniższym rysunku.

Rysunek 2. Kluczowe perspektywiczne obszary zastosowań ChatGPT



Źródło: J. Kocoń, I. Cichecki, O. Kaszyca et al., *ChatGPT: Jack of all trades, master of none*, „Information Fusion” 2023, vol. 99, DOI: 10.1016/j.inffus.2023.101861

W przyszłości kontrowersje związane z uznawaniem praw autorskich do dzieł tworzonych z zastosowaniem technologii AI mogą pogłębić się jeszcze dodatkowo, w sytuacji gdy inteligentne chatboty lub inne narzędzia oparte na technologii generatywnej sztucznej inteligencji zostaną zaimplementowane do autonomicznych robotów, także androidów, wyglądających jak ludzie. Jednak kluczowe jest to, aby w kwestii rozstrzygnięć dotyczących praw autorskich uwzględniać podstawowe zasady natury etycznej, tj. przede wszystkim relacje zachodzące między człowiekiem a narzędziem, którym jest robot lub inteligentny system informatyczny – narzędziem stworzonym przez człowieka i będącym jego własnością. Poza tym do etycznych działań zalicza się także te rozstrzygnięcia w kwestii uznania praw autorskich, które nie są sprzeczne z prawami tego rodzaju ustanowionymi wcześniej, nie pozabawiają określonych osób lub podmiotów tych praw w całości lub części

i/lub nie pozbawiają ich wynikających z tych praw określonych gratyfikacji, korzyści finansowych lub innych.

To, jak będą interpretowane w aspektach etycznych określone zastosowania sztucznej inteligencji, w tym sztucznej inteligencji generatywnej, niezależnie od zidentyfikowanego poziomu arcyzmu dzieł stworzonych przez inteligentne aplikacje wyposażone w technologię AI, zależy głównie od określenia relacji zachodzącej między człowiekiem występującym w roli twórcy zlecającego AI wykonanie określonych dzieł i tym, który odgrywa rolę twórcy technologii AI. W związku z tym technologia AI powinna być traktowana jako narzędzie w rękach człowieka – twórcy wykonującego określone dzieła w procesie tworzenia, w tym także w procesie kreatywnego tworzenia i procesie uznanym jako mniej lub bardziej etyczny. Ta kwestia powinna być uznana za fundamentalną i powinna być wystandaryzowana w skali globalnej. W przyszłości jednak sprawa ta może nadal się komplikować i uzyskiwać kolejne atrybuty niejednoznaczności, kontrowersyjności oraz stać się determinantem różnicującym możliwe paradygmaty uznanych za etyczne zalecanych i uregulowanych prawnie zastosowań technologii AI¹⁹. Stanie się tak, jeżeli w przyszłości będą tworzone (być może także produkowane seryjnie) wyposażone w generatywną sztuczną inteligencję autonomiczne roboty, w tym androidy wyglądające jak ludzie. Będą one zdolne do wykonywania różnego rodzaju prac twórczych, również takich, które do niedawna mogli wykonać tylko ludzie. Będą też wyposażone w coś, co obecnie jest określane jako „świadoma” sztuczna inteligencja, superinteligencja, osobliwość i/lub sztuczna inteligencja emocjonalna. Stanie się tak niezależnie od faktu, czy do tej pory w przyszłości zostanie już w pełni poznana istota ludzkiej świadomości czy też to, co obecnie powszechnie jest określane jako superinteligencja i/lub sztuczna inteligencja emocjonalna, zostanie uznane jedynie za generowaną przez inteligentną maszynę symulację ludzkiej świadomości i/lub inteligencji emocjonalnej. Jeżeli rozwój technologii i zastosowań AI będzie zmierzał także w tym kierunku, tj. w kierunku stworzenia inteligentnych, autonomicznych androidów, przypominających ludzi i wyposażonych

¹⁹ P. Henz. *Ethical and legal responsibility for artificial intelligence*, „Discover Artificial Intelligence” 2021, no. 1, s. 2–3.

w symulację „świadomej” sztucznej inteligencji i inteligencji emocjonalnej, to kwestia aspektów etycznych zapewne jeszcze niejednokrotnie powróci w dyskusjach i debatach dotyczących tej problematyki.

Innego rodzaju aspekty etyczne pojawiają się, gdy sztuczna inteligencja zostaje wykorzystana do działań przestępczych, jak np. do tworzenia nowych technik cyberprzestępczych i/lub hakerskich, polegających na włamywaniu się przez cyberprzestępców do systemów informatycznych innych podmiotów i instytucji celem dokonania kradzieży cennych danych, w tym wrażliwych danych osobowych, i/lub celem dokonania małwersacji finansowych²⁰. Tego rodzaju wykorzystanie technologii AI nie tylko jest nieetyczne, lecz jest także przestępstwem ściganym określonymi zapisami prawa²¹. Gdy natomiast ktoś tworzy nowe dzieło, m.in. przez włączenie do procesu jego tworzenia elementów ukradzionych z zastosowaniem wspomaganych technologią sztucznej inteligencji technik cyberprzestępczych z innych dzieł artystycznych lub innych chronionych prawami autorskimi, to czy powstałe w taki sposób nowe dzieło może być uznane za oryginalne? Do kogo będą należały prawa autorskie i w jakim zakresie zostaną złamane zdefiniowane prawnie zasady prawa autorskiego? Jest więc w tej kwestii wiele niuansów, które wymagają doprecyzowania i uregulowania prawnego.

Kwestia uregulowań prawnych rozwoju zastosowań technologii AI

Kwestia normatywów prawa dotyczącego zalecanych, możliwych prawnie zastosowań technologii sztucznej inteligencji, tj. doskonalenia np. zapisów AI Act, funkcjonującego w Unii Europejskiej, powinna być precyzyjnie definiowana, tak aby z jednej strony nie ograniczać rozwoju i zastosowań technologii AI, a z drugiej by nie stwarzać warunków do generowania

²⁰ G. Mazurek, K. Małagocka, *Perception of privacy and data protection in the context of the development of artificial intelligence*, „Journal of Management Analytics” 2019, no. 6(4), s. 347–348.

²¹ N. Bostrom, E. Yudkowsky, *The Cambridge hand-book of artificial intelligence*, w: *The ethics of artificial intelligence*, ed. K. Frankish, W. M. Ramsey, Cambridge University, 2014, s. 325–326.

działań nieetycznych, przestępczych, łamiących określone normy zachowań społecznych itd. Wspomniane doskonalenie normatywów prawa dotyczących zalecanych zastosowań technologii AI polega przede wszystkim na nowelizacji zapisów prawa celem dostosowywania ich treści do dokonującego się dynamicznie postępu technologicznego i stale obecnie się pojawiających nowych możliwych zastosowań tej technologii.

A więc z jednej strony rozwój technologii i zastosowań sztucznej inteligencji nie powinien być w prawie ograniczany, ponieważ takie ograniczanie skutkowałoby niższym poziomem innowacyjności, przedsiębiorczości gospodarczej i w konsekwencji także wolniejszym rozwojem gospodarczym²². Kraje, w których skala unormowania kwestii zastosowań technologii AI byłaby znacznie większa względem analogicznych regulacji w innych krajach, skazywałyby się na słabszą pozycję w dziedzinie postępu technologicznego i przełożenia tego postępu na rozwój gospodarczy²³. Z drugiej strony określone regulacje są jednak niezbędne, tak aby ograniczać skalę nieetycznych zastosowań sztucznej inteligencji i działalności przestępczej i/lub niezgodnej z obowiązującymi normami społecznymi. Nie można jednak na tak sformułowanej tezie poprzestawać. Niezbędne jest stałe monitorowanie rozwoju technologii AI, w tym generatywnej sztucznej inteligencji.

W przyszłości rozwój generatywnej sztucznej inteligencji może doprowadzić do powstania generalnej sztucznej inteligencji, tj. takiej jej formy, która z jednej strony dorówna lub przewyższy poziomem intelektualnego przetwarzania danych²⁴, informacji i wiedzy zdolności ludzkiego mózgu, a z drugiej strony uzyska zdolność autonomicznie realizowanego samodoskonalenia się, co może oznaczać możliwość wymknięcia się rozwojowi technologii AI spod kontroli człowieka. Być może dokonująca się obecnie rywalizacja czołowych firm działających w obszarze technologii informacyjno-komunikacyjnych

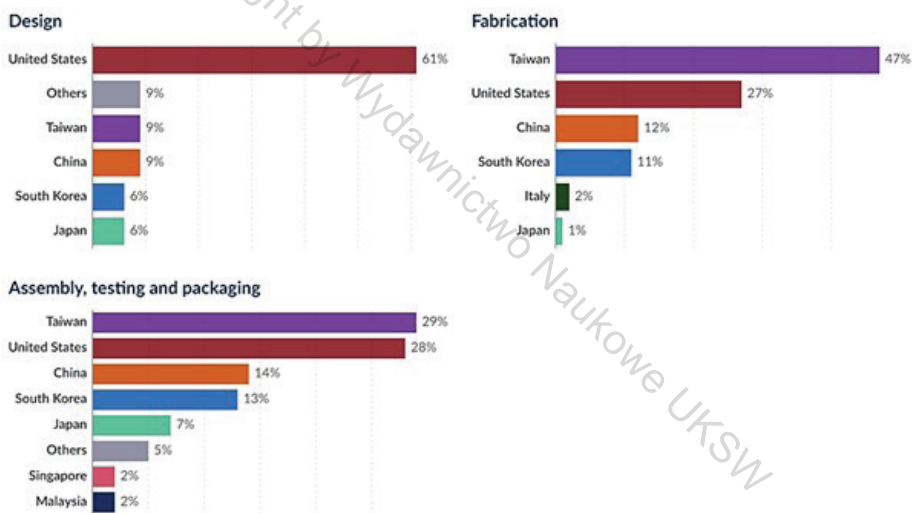
²² J. Furman, R. Seamans, *AI and the economy*, „Innovation Policy and the Economy” 2019, no. 19(1), s. 168–169.

²³ J. Mokyr, C. Vickers, N. L. Ziebarth, *The history of technological anxiety and the future of economic growth: Is this time different?*, „Journal of Economic Perspectives” 2015, no. 29(3), s. 34–35.

²⁴ J. Grzegorek, D. Prokopowicz, S. Gwoździewicz, *Zastosowanie zaawansowanych narzędzi przetwarzania danych w dobie cyfryzacji*, w: *Cyfryzacja w zarządzaniu*, red. A. Laskowska-Rutkowska, CeDeWu, Warszawa 2020, s. 101–102.

(ICT) w zakresie stworzenia coraz bardziej doskonałej generatywnej sztucznej inteligencji uitorowała drogę do zrealizowania się tego scenariusza w przyszłości. Z drugiej strony, w zakresie rozwoju technologii niezbędnych do rozwoju zastosowań AI, tj. np. technologii mikroprocesorów, przodują koncerny technologiczne mające swoje siedziby i zakłady produkcyjne głównie w kilku krajach rozwiniętych, w tym w USA i kilku krajach regionu Azji Południowo-Wschodniej. Unia Europejska występuje w tym rankingu na zdecydowanie słabszej pozycji. Wyjątkiem jest nieuwzględniona w poniższym zestawieniu Holandia, w której działa firma wytwarzająca podzespoły do mikroprocesorów produkowanych przez przodujące na tym rynku firmy z Tajwanu.

Wykres 4. Udział w rynku produkcji mikroprocesorów według etapów produkcji, 2021 r.



Łączny udział w rynku wszystkich firm z siedzibą w danym kraju. Podstawą są układy logiczne, takie jak procesory i procesory graficzne. Jednostki przetwarzania informacji w komputerach i innych urządzeniach elektronicznych.

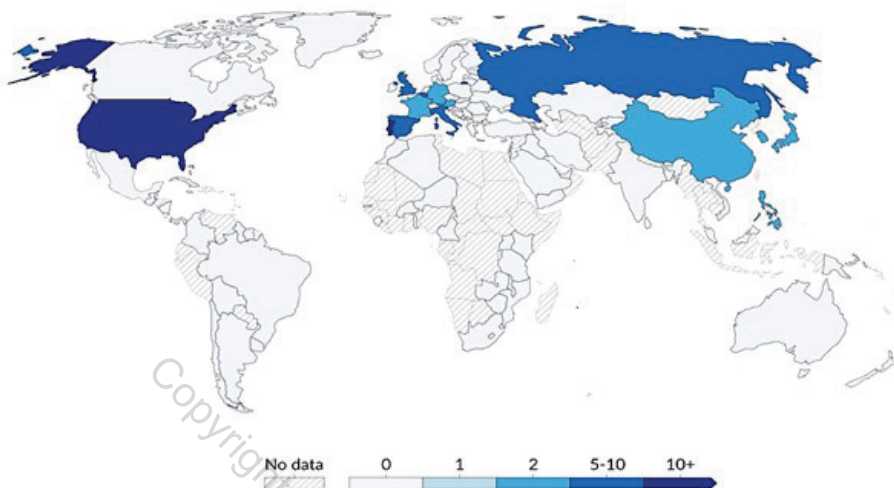
Źródło: Interactive Charts on Artificial Intelligence, <https://OurWorldInData.org/artificial-intelligence> (9.01.2024)

Warto odnotować, że w przyszłości powinna się zwiększyć skala unifikacji stosowanych zasad etycznego rozwoju technologii AI i jej zastosowań, jak również skala standaryzacji zapisów normatywów prawnych regulujących te kwestie. Zagadnienie to jest istotnym czynnikiem efektywnego i etycznego rozwoju technologii sztucznej inteligencji w warunkach międzynarodowej konkurencji firm i korporacji technologicznych rozwijających i wdrażających technologie AI w różnych dziedzinach zastosowań zwiększających efektywność procesów funkcjonujących w rozmaitych branżach i sektorach gospodarki. Firmy, w tym globalnie działające korporacje technologiczne przodujące w rozwoju nowych technologii informacyjnych ICT, internetowych i Przemysł 4.0/5.0²⁵, w tym także technologii AI, mają swoje siedziby w wielu krajach, także wysoko rozwiniętych i niektórych krajach rozwijających się, reprezentujących różne systemy polityczne, uwarunkowania struktur rynkowych oraz zróżnicowane podejście do kwestii etyki rozwoju i zastosowań technologii AI.

Już obecnie w zakresie etycznego stosowania technologii sztucznej inteligencji występują znaczące różnice, gdy porównuje się wdrażanie określonych technologii AI w biznesie np. w USA, Unii Europejskiej i Chinach. Występujące w tej kwestii różnice oraz wysoki poziom konkurencji, w tym rynkowej, biznesowej i technologicznej rywalizacji czołowych korporacji technologicznych, może powodować utrudnienia w sferze globalnej unifikacji zasad etycznego stosowania technologii AI. Poza tym konkurencja rynkowa, rywalizacja biznesowa i technologiczna może też być wzmocniana dodatkowo aspektami politycznymi, co jest związane z dużym wpływem działających międzynarodowo korporacji technologicznych na sferę polityki określonych krajów. Skalę międzynarodowego zróżnicowania unormowań prawnych odnoszących się do technologii AI ukazuje rysunek na kolejnej stronie.

²⁵ M. K. Such-Pyrgiel, D. Prokopowicz, A. Gołębiowska, *The postpandemic reality and the security of information technologies ICT, Big Data, Industry 4.0, social media portals and the Internet*, „Journal of Modern Science” 2022, vol. 49, no. 2, s. 17–18.

Rysunek 3. Projekty ustaw związanych ze sztuczną inteligencją, które weszły w życie w latach 2016–2023 r.



Projekty ustaw uchwalane przez krajowe organy ustawodawcze (np. Kongres, parlament) ze słowami kluczowymi „sztuczna inteligencja” (przetłumaczone na odpowiednie języki) w tytule lub treści ustawy. Na przykład *Ustawa o identyfikacji wyników generatywnych sieci przeciwstawnych* (*The Identifying Outputs of Generative Adversarial Networks Act*) została uchwalona przez Kongres USA.

Źródło: *Interactive Charts on Artificial Intelligence*, <https://OurWorldInData.org/artificial-intelligence> (9.01.2024)

W przyszłości najbezpieczniejszym dla człowieka rozwiązaniem w zakresie rozwoju technologii AI i jej zastosowań mogłoby być urzeczywistnienie się scenariusza, w którym synergia działań w obszarze ukierunkowanego doskonalenia technologii, jej zrównoważonego rozwoju²⁶, koegzystencji z człowiekiem oraz zastosowań zgodnych z etyką, normami społecznymi²⁷, moralnymi, normatywami prawa²⁸ i niegenerujących zastosowań antyspołecznych doprowadziły do stworzenia etycznej generatywnej sztucznej

²⁶ M. L. How, S. M. Cheah, Y. J. Chan, A. C. Khor, E. M. P. Say, *Artificial intelligence-enhanced decision support for informing global sustainable development: A human-centric AI-thinking approach*, „Information” 2020, no. 11(1), doi.org/10.3390/info11010039 39, s. 1–24.

²⁷ J. Baker-Brunnbauer, *Management perspective of ethics in artificial intelligence*, „AI and Ethics” 2021, no. 1(2), s. 176–177.

²⁸ D. Prokopowicz, *Opportunities and threats to the development of Artificial Intelligence applications and the need for normative regulation of this development*, „International Journal of Legal Studies” 2023, no. 2 (14), s. 112–113.

inteligencji²⁹, a w dalszej perspektywie być może także etycznej generalnej sztucznej inteligencji. Znaczenie kwestii koegzystencji z człowiekiem będzie w przyszłości rosło w związku z trwającymi już pracami nad stworzeniem na podobieństwo do ludzi inteligentnych maszyn w formie humanoidalnych, wyposażonych w generatywną sztuczną inteligencję autonomicznych androidów, które będą wyposażone w oprogramowanie etycznych, zgodnych z obowiązującymi normami zachowań społecznych, normatywami prawnymi itd.³⁰ W obszarze wspomnianej koegzystencji człowieka z inteligentnymi robotami w przyszłości, być może już nieodległej, zastosowania technologii AI będą rozwijane w formule tzw. *augmented intelligence*³¹, tj. z uwzględnieniem rozszerzania i wzmacniania zdolności poznawczych i kognitywistycznych człowieka. W związku z powyższym tego rodzaju zastosowania technologii AI, rozwijane zgodnie z zasadami etyki, określonymi normami społecznymi i normatywami prawa, będą mogły być określane jako etyczna rozszerzona inteligencja³².

Biorąc pod uwagę aspekty społeczne i etyczne właściwego stosowania technologii sztucznej inteligencji, warto dodać, że już obecnie duże znaczenie tej problematyki podkreśla także fakt zabrania głosu w odnośnym temacie przez papieża Franciszka. Otóż Ojciec Święty w Orędziu na 57. Światowy Dzień Pokoju, obchodzony 1 stycznia 2024 r., zachęca do uczynienia wszystkiego, aby postępy w rozwoju form sztucznej inteligencji służyły sprawie ludzkiego braterstwa i pokoju³³. Wspomniany komunikat papieski opublikowała Watykańska Dykasteria ds. Integralnego Rozwoju Człowieka³⁴.

²⁹ B. Mittelstadt, *Principles alone cannot guarantee ethical AI*, „Nature Machine Intelligence” 2019, no. 1(11), s. 504–505.

³⁰ B. C. Stahl, D. Schroeder, R. Rodrigues, *Ethics of Artificial Intelligence, Case Studies and Options for Addressing Ethical Challenges*, Springer, Bath 2023, s. 111.

³¹ I. Kushchu, I. Shnurenko, T. Murovana, *Artificial Intelligence Media and Information Literacy, Human Rights and Freedom of Expression*, UNESCO, Moscow 2020, s. 4.

³² M. Nowakowski, *O moralnej odpowiedzialności HAL-a 9000, czyli etyka sztucznej inteligencji w praktyce: czy potrzebujemy definicji sztucznej inteligencji?*, „Kwartalnik Naukowy Prawo Mediów Elektronicznych” 2022, nr 1, DOI: 10.34616/144181.

³³ K. Dudek, *Papież: sztuczna inteligencja może być szansą, jak i zagrożeniem*, 14.12.2023, <https://www.vaticannews.va/pl/papiez/news/2023-12/papiez-sztuczna-inteligencja-moze-byc-szansa-jak-i-zagrozeniem.html>.

³⁴ *Papież Franciszek zabrał głos w sprawie sztucznej inteligencji*, 8.08.2023, <https://www.rp.pl/kosciol/art38919051-papiez-franciszek-zabral-glos-w-sprawie-sztucznej-inteligencji>.

Papież Franciszek wskazuje w nim, że nauka i technologia są niezwykleymi wytworami potencjału twórczego ludzkiego rozumu. Dokonujący się postęp w tej dziedzinie może z jednej strony prowadzić do udoskonalenia człowieka i pozytywnej przemiany świata, ale z drugiej – do powstawania zagrożeń dla ludzkiego przetrwania i środowiska naturalnego³⁵.

Papież Franciszek wskazuje, że rozwój technologii AI powinien się dokonywać zgodnie z zasadami *algor-etyki*, tj. etycznego rozwoju algorytmów, które rozwijałyby się w sposób ukierunkowany z uwzględnieniem określonych ludzkich systemów wartości, bazujących m.in. na etyce³⁶, powszechnie przyjętych i społecznie akceptowalnych standardach poprawnego funkcjonowania, normatywach prawnych³⁷, wartościach wynikających z religii itd. Aspekty etyczne powinny być uwzględniane na każdym etapie rozwoju technologii AI i jej zastosowań, tj. na etapie badań, testowania, projektowania, produkcji, dystrybucji, wprowadzenia na rynek i zastosowania do określonych działalności.

Zakończenie

W niniejszym rozdziale przedstawiono wybrane istotne kwestie dotyczące rozwoju i zastosowań technologii sztucznej inteligencji z uwzględnieniem aspektów etycznych i społecznych. Głównym celem tekstu było wskazanie na rosnącą skalę rozwoju i zastosowań technologii generatywnej sztucznej inteligencji w różnych dziedzinach działalności podmiotów gospodarczych, w tym w zastosowaniach badawczych, analitycznych, biznesowych, marketingowych, logistycznych, cyberbezpieczeństwa i innych. Od czasu pandemii COVID-19 nastąpiło przyśpieszenie rozwoju technologii sztucznej inteligencji, w tym technologii algorytmów generatywnych, uczących się maszyn

³⁵ *Sztuczna inteligencja. Papież zabiera głos*, 14.12.2023, <https://dorzczy.pl/swiat/525402/papiez-zabral-glos-ws-sztucznej-inteligencji.html>.

³⁶ R. Nath, V. Sahu, *The problem of machine ethics in artificial intelligence*, „Artificial Intelligence & Society” 2020, no. 35, s. 106–107.

³⁷ T. Tzimas, *Legal and Ethical Challenges Artificial Intelligence from an International Law Perspective*, Springer Nature Switzerland AG, Switzerland, 2021, s. 41.

i głębokiego uczenia³⁸. Przyspieszeniu uległ proces uczenia algorytmów generatywnych wykonywania określonych czynności, które do niedawna wykonywał tylko człowiek.

Z rozwojem tym wiążą się zarówno aspekty pozytywne, jak i negatywne, tj. i szanse rozwojowe, i zagrożenia dla rozwoju ludzkiej cywilizacji. Z jednej strony rozwój generatywnej sztucznej inteligencji jest swego rodzaju kolejnym etapem dokonującego się już od epoki pierwszej rewolucji przemysłowej postępu technologicznego, polegającego na wzroście skali uprzedmiotowienia pracy i automatyzacji procesów wytwórczych. Z drugiej strony, biorąc pod uwagę udostępniony w otwartym dostępie w internecie ChatGPT, rozwój sztucznej inteligencji generuje także wiele zagrożeń związanych z potencjalnie szybkim rozwojem dezinformacji w mediach społecznościowych, nieprzestrzeganiem praw autorskich, zmniejszającą się możliwością identyfikacji autorstwa dzieł stworzonych przez sztuczną inteligencję, zastąpieniem technologią człowieka na wielu stanowiskach pracy oraz wykorzystaniem generatywnej sztucznej inteligencji przez hakerów i cyberprzestępców do doskonalenia hakerskich technik nieuprawnionego włamywania się do systemów informatycznych innych instytucji, tworzeniem nowych rodzajów wirusów komputerowych, rozprzestrzenianiem przez narzędzia zdalnej komunikacji internetowej złośliwego oprogramowania i infekowania tym oprogramowaniem coraz większej liczby komputerów, laptopów, smartfonów oraz innych urządzeń mobilnych podłączonych do internetu.

W związku z tym w kolejnych latach będzie się zapewne dokonywać kolejny etap swoistego wyścigu zbrojeń między hakerami i cyberprzestępcami a informatykami, których zadaniem jest doskonalenie systemów cyberbezpieczeństwa³⁹. Cechą charakterystyczną tego nowego etapu wspomnianej nieformalnej rywalizacji jest wykorzystanie sztucznej inteligencji po dwóch

³⁸ M. Matosek, D. Prokopowicz, A. Gołębiowska, *Growing importance of digitalization of remote communication processes and the internetization of economic processes and the impact of the SARS-CoV-2 (Covid-19) coronavirus pandemic on the economy*, w: *Socio-economic and legal dimensions of digital transformation. Selected contexts*, ed. A. Gołębiowska, M. Such-Pyrgiel, Wydawnictwo SGSP, Warszawa 2021, s. 234–235.

³⁹ D. Prokopowicz, A. Gołębiowska, M. Such-Pyrgiel, *The role of Big Data and Data Science in the context of information security and cybersecurity*, „Journal of Modern Science” 2023, vol. 53, no. 4, s. 14–15.

stronach barykady. Poza tym z rozwojem tym wiąże się wiele ryzyk i zagrożeń, których skalę można znacznie ograniczyć, jeżeli na każdym etapie procesu tworzenia, rozwoju i wdrażania technologii AI zostaną uwzględnione także kwestie etyczne i aspekty prospołeczne⁴⁰, a cały wspomniany proces zostanie normatywnie uregulowany, co powinno w dużym stopniu zmniejszyć skalę prawdopodobieństwa wymknięcia się tego rozwoju spod kontroli człowieka. W związku z powyższym, z dynamicznie dokonującym się obecnie rozwojem zastosowań biznesowych technologii generatywnej sztucznej inteligencji wiążą się liczne aspekty natury społecznej i etycznej. Niezbędne jest doskonalenie regulacji prawnych normujących kwestię rozwoju i zastosowań technologii sztucznej inteligencji, aby ten rozwój nie wymknął się spod kontroli człowieka i nie generował nowych zagrożeń i ryzyk w obszarach ludzkiej działalności. Regulacje prawne dotyczące zastosowań technologii generatywnej sztucznej inteligencji powinny uwzględniać aspekty społeczne i etyczne.

W kontekście powyższych rozważań w niniejszym rozdziale zaprezentowano sformułowaną przez autora teorię równowagi w kwestii przyznawania decyzyjności inteligentnym systemom informatycznym, maszyn i autonomicznych robotów wyposażonych w sztuczną inteligencję. Teoria ta polega na odpowiednio symetrycznym skorelowaniu wspomnianej decyzyjności z przeniesionym na smarttechnologie i systemy informatyczne zakresem odpowiedzialności za realizowane zleczone zadania oraz na uwzględnieniu etycznego stosowania⁴¹ technologii AI i kontroli człowieka dotyczącej poziomu rzetelności etycznego rozwoju zastosowań AI. Wspomniana kontrola powinna uwzględniać także kwestie etyczne i społeczne zgodnego z potrzebami ludzi wykorzystania technologii AI w jej różnych zastosowaniach.

W kontekście aspektów etycznych omówiono kwestię praw autorских do dzieł tworzonych z zastosowaniem generatywnej sztucznej inteligencji. W niniejszym rozdziale wskazano także na konieczność ciągłego

⁴⁰ T. Turan, G. Turan, E. U. Kucuksille, *Ethics of Artificial Intelligence: Impact on Society*, „The Journal of Graduate School of Natural and Applied Sciences of Mehmet Akif Ersoy University” 2022, no. 13(2), s. 294–295.

⁴¹ A. B. Brendel, M. Mirbabaie, T. B. Lembcke, L. Hofeditz, *Ethical management of artificial intelligence*, „Sustainability” 2021, no. 13(4), DOI: 10.3390/su13041974, s. 1–18.

dostosowywania uregulowań prawnych do dynamicznie rozwijającej się technologii sztucznej inteligencji. Wspomniane uregulowania nie powinny prowadzić do ograniczenia skali tego rozwoju i tym samym do spowolnienia dokonującego się postępu technologicznego, lecz do ograniczenia skali negatywnych skutków zastosowań nowych technologii. Proces nowelizacji tych regulacji powinien być prowadzony na tyle sprawnie, aby nie pozostawał w tyle za wspomnianym rozwojem owych technologii, w tym technologii generatywnej sztucznej inteligencji, jak również powinien uwzględniać kwestie etyczne i społeczne.

Bibliografia

- Annual scholarly publications on AI, by field of study*, <https://ourworldindata.org/grapher/number-ai-publications-by-field> (20.12.2023).
- Baker-Brunnbauer J., *Management perspective of ethics in artificial intelligence*, „AI and Ethics” 2021, no. 1(2), s. 173–181.
- Bélisle Pilon J. C., Monteferrante E., Roy M. C., Couture V., *Artificial intelligence ethics has a black box problem*, „AI & Society” 2022, s. 1–16.
- Bill Gates o rewolucyjnej technologii*. „Zmieni sposób, w jaki ludzie pracują, uczą się, podróżują”, 22.03.2023, <https://tvn24.pl/biznes/ze-swiata/chat-gpt-bill-gates-o-rewolucyjnej-technologie-zmieni-sposob-w-jaki-ludzie-pracuja-ucza-sie-podrozuja-6854903>.
- Bostrom N., Yudkowsky E., *The Cambridge hand-book of artificial intelligence*, w: *The ethics of artificial intelligence*, ed. K. Frankish, W. M. Ramsey, Cambridge University, 2014, s. 316–334.
- Brendel A. B., Mirbabaie M., Lembcke T. B., Hofeditz L., *Ethical management of artificial intelligence*, „Sustainability” 2021, no. 13(4), DOI: 10.3390/su13041974, s. 1–18.
- Cath C., *Governing Artificial Intelligence: Ethical, legal, and technical opportunities and challenges*, Royal Society, London 2018.
- Day B., *Basic Concepts of Big Data and Artificial Intelligence in Managing Climate Data*, <https://www.climate4life.info/2018/12/kompasi-dasar-big-data-dan-artificial-intelligence-dalam-pengelolaan-data-iklim.html> (4.11.2023).
- Dmowski A., Sarnowski J., Prokopowicz D., *Podstawy finansów i bankowości*, Difin, Warszawa 2005.
- Dudek K., *Papież: sztuczna inteligencja może być szansą, jak i zagrożeniem*, 14.12.2023, <https://www.vaticannews.va/pl/papiez/news/2023-12/papiez-sztuczna-inteligencja-moze-byc-szansa-jak-i-zagrozeniem.html>.
- Furman J., Seamans R., *AI and the economy*, „Innovation policy and the economy” 2019, no. 19(1), s. 161–191.

- Gołębiowska A., Prokopowicz D., *Growing importance of ICT, Industry 4.0 and Big Data Analytics as key determinants of the development of The Financial Industry 4.0*, w: *Socio-economic and legal dimensions of digital transformation. Selected contexts*, ed. A. Gołębiowska, M. Such-Pyrgiel, Wydawnictwo SGSP, Warszawa 2021, s. 35–56.
- Grzegorek J., Prokopowicz D., Gwoździewicz S., *Zastosowanie zaawansowanych narzędzi przetwarzania danych w dobie cyfryzacji*, w: *Cyfryzacja w zarządzaniu*, red. A. Laskowska-Rutkowska, CeDeWu, Warszawa 2020, s. 93–128.
- Gwoździewicz S., Prokopowicz D., *Determinants of electronic data interchange security in the context of Big Data and cloud computing technology development. Sentiment analysis conducted for cybercrime issues occurred in the period from May 2017 to February 2019*, „International Journal of New Economics and Social Sciences” 2022, vol. 15, no. 1, s. 77–109.
- Henz P. *Ethical and legal responsibility for artificial intelligence*, „Discover Artificial Intelligence” 2021, no. 1, s. 1–5.
- How M. L., Cheah S. M., Chan Y. J., Khor A. C., Say E. M. P., *Artificial intelligence-enhanced decision support for informing global sustainable development: A human-centric AI-thinking approach*, „Information” 2020, no. 11(1), doi.org/10.3390/info11010039 39, s. 1–24.
- Huyue-Zhang A., *China generative AI regulation support growth*, październik 2023, <https://project-syndicate.org/commentary/china-generative-ai-regulation-support-growth-by-angela-huyue-zhang-2023-10>.
- Interactive Charts on Artificial Intelligence*, <https://OurWorldInData.org/artificial-intelligence> (9.01.2024).
- Kocoń J., Cichecki I., Kaszyca O. et al., *ChatGPT: Jack of all trades, master of none*, „Information Fusion” 2023, vol. 99, DOI: 10.1016/j.inffus.2023.101861.
- Kuchta A., *Sztuczna inteligencja przyspiesza powstanie leków*, 16.10.2023, <https://ai.infor.pl/sztuczna-inteligencja/6338561,sztuczna-inteligencja-przyspiesza-powstanie-lekow.html>.
- Kushchu I., Shnurenko I., Murovana T., *Artificial Intelligence Media and Information Literacy, Human Rights and Freedom of Expression*, UNESCO, Moscow 2020.
- Lesandrini J., Idris M. Y., Reis D. S., *The Ethics of Artificial Intelligence and Machine Learning*, „Journal of Radiology Nursing” 2023, no. 42(3), s. 265–266.
- Matosek M., Prokopowicz D., Gołębiowska A., *Growing importance of digitalization of remote communication processes and the internetization of economic processes and the impact of the SARS-CoV-2 (Covid-19) coronavirus pandemic on the economy*, w: *Socio-economic and legal dimensions of digital transformation. Selected contexts*, ed. A. Gołębiowska, M. Such-Pyrgiel, Wydawnictwo SGSP, Warszawa 2021, s. 221–250.
- Mazurek G., Małagocka K., *Perception of privacy and data protection in the context of the development of artificial intelligence*, „Journal of Management Analytics” 2019, no. 6(4), s. 344–364.
- Mittelstadt B., *Principles alone cannot guarantee ethical AI*, „Nature Machine Intelligence” 2019, no. 1(11), s. 501–507.

- Mokyr J., Vickers C., Ziebarth N. L., *The history of technological anxiety and the future of economic growth: Is this time different?*, „Journal of Economic Perspectives” 2015, no. 29(3), s. 31–50.
- Murphy K., Di Ruggiero E., Upshur R., Willison D. J., Malhotra N., Cai J. C., Gibson J., *Artificial intelligence for good health: a scoping review of the ethics literature*, „BMC Medical Ethics” 2021, no. 22(1), s. 1–17.
- Nath R., Sahu V., *The problem of machine ethics in artificial intelligence*, „Artificial Intelligence & Society” 2020, no. 35, s. 103–111.
- Nowakowski M., *O moralnej odpowiedzialności HAL-a 9000, czyli etyka sztucznej inteligencji w praktyce: czy potrzebujemy definicji sztucznej inteligencji?*, „Kwartalnik Naukowy Prawo Mediów Elektronicznych” 2022, nr 1, DOI: 10.34616/144181.
- Olewiński A., *AI w białym fartuchu. Sztuczna inteligencja i nowe leki. Jak jest i jak będzie wykorzystywana w Polsce, Europie i na świecie?*, 24.07.2023, <https://aptekarski.com/artykul/ai-w-bialym-fartuchu-sztuczna-inteligencja-i-nowe-leki-jak-jest-i-jak-bedzie-wykorzystywana-w-polsce-europie-i-na-swiecie>.
- Papież Franciszek zabrał głos w sprawie sztucznej inteligencji*, 8.08.2023, <https://www.rp.pl/kosciol/art38919051-papiez-franciszek-zabral-glos-w-sprawie-sztucznej-inteligencji>.
- Pizło W., Kulykovets O., Prokopowicz D., Mazurkiewicz-Pizło A., Kałowski A., Paprocka M. W., Stawicka E., Skarzyńska E., *The importance of Big Data Analytics technology in business management*, „Cybersecurity and Law” 2023, no. 2 (10), s. 270–282.
- Prokopowicz D., *Opportunities and threats to the development of Artificial Intelligence applications and the need for normative regulation of this development*, „International Journal of Legal Studies” 2023, no. 2 (14), s. 95–130.
- Prokopowicz D., Gołębiowska A., Such-Pyrgiel M., *The role of Big Data and Data Science in the context of information security and cybersecurity*, „Journal of Modern Science” 2023, vol. 53, no. 4, s. 9–42.
- Robot ethics 2.0: From autonomous cars to artificial intelligence*, ed. P. Lin, K. Abney, R. Jenkins, Oxford University Press, 2017.
- Stahl B. C., Schroeder D., Rodrigues R., *Ethics of Artificial Intelligence, Case Studies and Options for Addressing Ethical Challenges*, Springer, Bath 2023.
- Stylec-Szromek P., *Sztuczna inteligencja – prawo, odpowiedzialność, etyka*, „Zeszyty Naukowe Politechniki Śląskiej” 2018, z. 123, s. 501–509.
- Such-Pyrgiel M. K., Prokopowicz D., Gołębiowska A., *The postpandemic reality and the security of information technologies ICT, Big Data, Industry 4.0, social media portals and the Internet*, „Journal of Modern Science” 2022, vol. 49, no. 2, s. 10–43.
- Sztuczna inteligencja. Papież zabiera głos*, 14.12.2023, <https://dorzczy.pl/swiat/525402/papiez-zabral-glos-ws-sztucznej-inteligencji.html>.
- Turan T., Turan G., Kucuksille E. U., *Ethics of Artificial Intelligence: Impact on Society*, „The Journal of Graduate School of Natural and Applied Sciences of Mehmet Akif Ersoy University” 2022, no. 13(2), s. 292–299.
- Tzimas T., *Legal and Ethical Challenges Artificial Intelligence from an International Law Perspective*, Springer Nature Switzerland AG, 2021.
- Yerukala M., *Artificial Intelligence (AI) Projects*, <https://mindmajix.com/artificial-intelligence-projects> (17.08.2023).

- Yerukala M., *Artificial Intelligence Tutorial*, <https://mindmajix.com/artificial-intelligence-tutorial> (16.12.2023).
- Yusep Multana H., *Ethics and Law for the Use and Development of Artificial Intelligence Technology*, „East Asian Journal of Multidisciplinary Research” 2023, vol. 2, no. 11, s. 4519–4532.

Development and applications of artificial intelligence technologies with ethical and social considerations

Summary

This chapter describes selected relevant issues concerning the development and applications of artificial intelligence technology, taking into account ethical and social aspects. With the dynamic development of business applications of generative artificial intelligence technology that is currently taking place, there are numerous social and ethical aspects involved. It is necessary to improve the legal regulations that normalize the development and applications of artificial intelligence technology, so that this development does not escape human control and does not generate new threats and risks in areas of human activity. Legal regulations on the applications of generative artificial intelligence technology should take into account social and ethical aspects. In the context of the above considerations, this chapter presents the equilibrium theory formulated by the author on the issue of granting decision-making authority for intelligent information systems. In the context of ethical aspects, the issue of copyright for works created using generative artificial intelligence is discussed. The voice of Pope Francis on the issue of the future development of artificial intelligence applications is presented, and the need to constantly adapt legal regulations to the rapidly developing artificial intelligence technology is also pointed out.

Keywords: artificial intelligence, generative artificial intelligence, ethics of artificial intelligence, social aspects of artificial intelligence development, risks of artificial intelligence development, new technologies, copyright vs. artificial intelligence technology

**II. Wyzwania związane
z zastosowaniem
sztucznej inteligencji
w różnych sferach życia społecznego**

Copyright by Wydawnictwo Naukowe UKSW

Sztuczny nauczyciel – prawdziwe konsekwencje wychowawcze

Wstęp

Wpływ nauki i techniki na człowieka, a jeszcze silniej na człowieczeństwo, staje się głównym źródłem współczesnych zmian. Nie zawsze oznaczają one rozwój, dlatego człowiek od początku swego istnienia podejmuje wysiłki mające na celu przewidzenie i zrozumienie tego, co potencjalnie może przynieść przyszłość. Ocena nauki i techniki nie za każdym razem w tych prognozach jest pozytywna. W 1749 r. Akademia w Dijon ogłosiła konkurs, w ramach którego należało odpowiedzieć na pytanie: Czy postęp nauki i sztuki przyczynia się do poprawy obyczajów? Zwycięzca konkursu, Jean-Jacques Rousseau, w eseju *Discours sur les sciences et les arts* (1750 r.), przedstawił w dwóch częściach diatrybę² przeciwko nauce i sztuce. Według autora psują one obyczaje i oddalają człowieka od doskonałości moralnej. Co więcej, rozwój nauki spowodował, że ludzie przestali pamiętać o tym, co jest istotne, a zajmując się błahostkami, zapominają o wolności i tym samym stają się coraz bardziej zniewoleni. Pomnożenie wygod życia sprawiło,

¹ Dr hab. Sylwia Jaskuła, prof. AŁ, Wydział Nauk Społecznych i Humanistycznych, Akademia Łomżyńska.

² Diatryba – przemówienie albo utwór literacki wyrażający protest lub krytykę, *Słownik języka polskiego*, PWN, <https://sjp.pwn.pl/szukaj/diatryba.html>.

że zniknęły liczne wartości, a to z kolei wpłynęło negatywnie na zachowanie przez jednostki podstawowych cech moralnych³.

Nowe rozwiązania technologiczne, mające przynosić ludzkości zażyteczności, nie zawsze są wykorzystywane zgodnie z pierwotnymi założeniami ich zastosowania, a skutki ich wdrażania są często nieprzewidywalne. Dzieje się tak głównie dlatego, że technologie są przyjmowane w nowych kontekstach w stosunku do czasu ich projektowania, dodatkowo często są wykorzystywane w sposób, który odbiega od założeń ich użycia przewidzianych przez twórców. Internet, opracowany w celu ułatwienia przesyłania informacji między sieciami komputerowymi, stał się kluczowym narzędziem handlu, wprowadzając nieoczekiwane możliwości w zakresie przestępczości i oszustw finansowych. Platformy mediów społecznościowych, takie jak Facebook i Twitter, zaprojektowane w celu łączenia przyjaciół i rodzin, stały się centralnymi punktami wpływów politycznych. Jednym z kluczowych wyzwań w obszarze tworzenia nowych technologii jest zatem rozpoznanie, a nawet przewidzenie niektórych z tych nieoczekiwanych konsekwencji oraz zapewnienie technologom mechanizmów pozwalających na przeanalizowanie potencjalnego wpływu wytworów ich pracy, a także możliwych ścieżek prowadzących do różnych skutków.

Założenie to jest szczególnie trudne w przypadku wykorzystania i zastosowania sztucznej inteligencji, której drogi rozwoju na dzisiejszym etapie jej planowania są bardzo trudne do przewidzenia. Próby wdrażania jej w coraz większą liczbę obszarów, w których możliwe jest wsparcie przez nią działań podejmowanych przez człowieka, są jednocześnie potencjalnymi sferami zagrożeń. Tym samym istnieją dziedziny, w których sztuczna inteligencja nie znajdzie obecnie zastosowania ze względu na przewagę niebezpieczeństw nad pozytywnymi, ale istnieją również takie przestrzenie, w których nigdy nie powinna dominować. Jednym z takich obszarów jest niewątpliwie wychowanie. Jego złożoność i wieloaspektowość, a przede wszystkim konieczność udziału w nim podmiotów, które przygotowują jednostki do życia w społeczeństwie oraz rozwijają ich potencjały, wyklucza oddziaływanie systemów

³ A. Zwoliński, *Cyberpułapka. Jak internet, sztuczna inteligencja i wirtualna rzeczywistość skutecznie niszczą cywilizację*, Fronda, Warszawa 2023, s. 7.

pozbawionych cech podmiotowych, a za takie należy uznać sztuczną inteligencję, z wszelkimi jej antropomorficznymi⁴ odmianami.

Romantyzm technologiczny

Postęp technologii i jej coraz znaczniejszy wpływ na rozwój ludzkości wygenerowały pytania o jej oddziaływanie na procesy ewolucji pojedynczego człowieka, na efekty działań grup społecznych, ale również na transformacje całego świata. Kwestie te stają się coraz bardziej istotne wobec rozpowszechniania się technologii, które wraz ze swoim rozwojem stają się bardziej skomplikowane i mocniej wzajemnie powiązane. Ekspansja sztucznej inteligencji nadała technice nowe cechy, tj. niezrozumienie, niekontrolowalność, nieograniczoność czy nieprzewidywalność. Samochody, samoloty, urządzenia medyczne czy transakcje finansowe opierają się na większej niż kiedykolwiek wcześniej liczbie oprogramowania komputerowego wspieranego sztuczną inteligencją. Ich mnogość sprawia, że są coraz trudniejsze do zinterpretowania, ale także w niektórych przypadkach trudniejsze do nadzorowania, a tym samym coraz odleglejsze od człowieka.

Sam proces przetwarzania informacji opiera się współcześnie w dużej mierze na technologiach cyfrowych i sztucznej inteligencji, a zatem wiąże się z mniejszą niż kiedykolwiek wcześniej obecnością człowieka. Może to prowadzić do tworzenia informacji nie tylko nieprzydatnych w kategoriach ludzkich wartości, ale przede wszystkim mogą powstawać dane, których człowiek nie będzie w stanie zidentyfikować, a co gorsza – zrozumieć i wykorzystać. Dane te mogą z kolei być nadal przetwarzane przez sztuczną inteligencję, prowadząc do powstania jeszcze bardziej skomplikowanych i niezrozumiałych dla człowieka struktur, a nawet całych ich systemów. Zmiany te i wiele innych będących ich konsekwencją będą coraz silniej wpływać na osobowość i cechy samego człowieka, pogłębiając redukcyjne komponenty transformacji

⁴ Antropomorfizacja – przenośnia polegająca na przypisywaniu cech ludzkich zwierzętom, roślinom, przedmiotom, zjawiskom lub pojęciom abstrakcyjnym, *Słownik języka polskiego*, PWN, <https://sjp.pwn.pl/szukaj/antropomorfizacja.html>.

w relacjach między nim i jego rozszerzoną globalną rzeczywistością⁵. Przemiany tej nie unikniemy, nawet mając świadomość, że nie zawsze jest ona związana z jego autentycznym rozwojem. Ewolucja jest immamentnym aspektem historii cywilizacji.

Człowiek nie może żyć statycznie, ponieważ wewnętrzne sprzeczności popychają go ku poszukiwaniu równowagi, nowej harmonii w miejsce utraconej zwierzęcej harmonii z przyrodą. Zaspokoiwszy swoje potrzeby zwierzęce, odczuwa nacisk potrzeb ludzkich. Podczas gdy ciało mówi mu, co ma jeść, a czego unikać, sumienie powinno mu powiedzieć, jakie potrzeby ma kultywować i zaspokajać, a jakie stłumić i pokonać [...]. Wszystkie namiętności i pragnienia człowieka są próbami znalezienia odpowiedzi na jego egzystencję albo, mówiąc inaczej, są próbą uniknięcia szaleństwa⁶.

Czasy rozwoju technologicznego to także okres nasilenia polityki technologicznej, której celem jest zapewnienie równowagi między wprowadzaniem nowych rozwiązań technologicznych a zapewnieniem bezpieczeństwa i zaufania w cyfrowym świecie. Nowa „technologia intelektualna”⁷ jako efekt przemian społecznych opisanych przez Daniela Bella w 1963 r. jest coraz bardziej zaawansowana i jednocześnie skomplikowana. Zachowaniu równowagi między możliwymi drogami jej rozwoju a sposobami jej wykorzystania powinno towarzyszyć coraz pełniejsze rozumienie, w jaki sposób można ją najlepiej zastosować, ale przede wszystkim kontrolować, a nawet ograniczyć. Obecnie prowadzonych jest coraz więcej badań, w efekcie których możliwe będzie doskonalsze nadzorowanie sztucznej inteligencji, choć wciąż

⁵ Por. S. Jaskuła, *Jak sztuczna inteligencja zmienia świat*, w: *Świat w czasie przełomu. Religijne mosty nadziei*, red. P. Walewski, Bernardinum, Pelplin 2023.

⁶ E. Fromm, *Zdrowe społeczeństwo*, tłum. A. Tanalska-Dulęba, Vis-à-Vis Etiuda, Warszawa 1996, s. 42.

⁷ Bell definiuje „pięć wymiarów społeczeństwa postindustrialnego” obejmującego: (1) stworzenie gospodarki opartej na usługach; (2) prymat klasy profesjonalnej i technicznej; (3) główne znaczenie wiedzy teoretycznej jako źródła innowacji i formułowania polityki w społeczeństwie; (4) możliwość samonapędzającego się wzrostu technologicznego; (5) stworzenie nowej „technologii intelektualnej”. D. Bell, *The Measurement of Knowledge and Technology*, w: *Indicators of Social Change*, ed. E. Sheldon, W. Moore, Russell Sage Foundation, New York 1968, s. 152–153.

przybywa sceptyków co do skuteczności tej kontroli. 6 października 2022 r. ukazał się raport z badań przeprowadzonych przez naukowców z Instytutu Maxa Plancka, z którego to raportu wynika, że ludzie nie będą zdolni do kontrolowania superinteligentnej SI⁸. W badaniach nad nadzorowaniem sztucznej inteligencji w procesach podejmowania decyzji niezależnych od ludzi naukowcy wykorzystali tzw. problem stopu Alana Turinga, analizujący, czy program, który realizuje zadany algorytm, zakończy pracę, czy będzie ją wykonywał w nieskończoność. Testy przeprowadzane na sztucznej inteligencji pokazały, że zastosowanie tzw. algorytmu powstrzymania, którego celem jest regulacja zachowań maszyny w przypadku działań szkodliwych dla ludzi, jest niemożliwe. Tym samym naukowcy sformułowali konkluzję, odnosząc je do wysokiego ryzyka, iż w procesie uczenia się sztuczna inteligencja może dojść do własnych, niezależnych wniosków. Dlatego jeśli ludzkość może funkcjonować bez AI, to nie mając nad nią pełnej kontroli, nie powinna jej rozwijać.

Równoległe do toczących się badań trwają prace nad opracowaniem właściwych regulacji prawnych oraz etycznych pozwalających zachować równowagę między możliwościami współczesnych technologii a potrzebami i umiejętnościami człowieka⁹. Otwarte pozostaje pytanie o to, czy jesteśmy w stanie opracować takie przepisy, które w pełni zapewnią bezpieczeństwo. Musiałoby to być prawo bezwzględnie podporządkowujące roboty ludziom i w ten sposób zapewniające im bezpieczeństwo i wyłączenie z czerpania korzyści.

Problem zasad moralnych pojawił się znacznie wcześniej niż realne możliwości produkcji sztucznej inteligencji. Powstałe w roku 1942 cztery prawa robotów Isaaca Asimova, regulujące stosunki między myślącymi maszynami a ludźmi i przedstawione w opowiadaniach *Zabawa w berka* (*Runaround*) oraz *Roboty i imperium* (*Robots and Empire*), zakładały:

1. Robot nie może skrzywdzić ludzkości ani przez wstrzymanie się od działania pozwolić na wyrządzenie jej krzywdy¹⁰.

⁸ A. Andreu, J. Ravindran, *Ludzie nie będą zdolni do kontrolowania superinteligentnej AI. To wnioski z badania*, 6.10.2022, <https://businessinsider.com.pl/technologie/nowe-technologie/naukowcy-sztuczna-inteligencja-bedzie-nie-do-opanowania/qewm60w> (21.12.2023).

⁹ D. Gunning, D. W. Aha, *DARPA's Explainable Artificial Intelligence Program*, „AI Magazine” 2019, 40 (2), s. 44–58.

¹⁰ I. Asimov, *Zabawa w berka*, w: tegoż, *Ja, robot*, tłum. Z. A. Królicki, Rebis, Poznań 1980.

2. Robot nie może zrobić niczego, co mogłoby zaszkodzić człowiekowi, ani też, przez beczynność, nie może pozwalać na to, by człowiekowi stała się krzywda.
3. Robot musi być posłuszny człowiekowi, z wyjątkiem sytuacji, kiedy byłoby to sprzeczne z Pierwszym Prawem.
4. Robot musi chronić siebie, o ile tylko nie stoi to w sprzeczności z Pierwszym lub Drugim Prawem¹¹.

Kodeks Asimova został skrytykowany. Sam twórca w swojej powieści *Nagie słońce* wskazuje przez głównego bohatera Elijaha Baleya, że roboty mogą „nieświadomie” złamać każde z praw. Powtórzył drugie prawo w następujący sposób: „Robot nie może zrobić niczego, co, *według jego wiedzy*, mogłoby zaszkodzić człowiekowi, ani też, poprzez beczynność, *świadomie* nie może pozwalać na to, by człowiekowi stała się krzywda”. Ta zmiana brzmienia jasno pokazuje, że roboty mogą się stać narzędziami morderstwa, pod warunkiem że nie są świadome natury swoich zadań. Wielokrotnie potem wskazywano, że system moralny Asimova nie może działać poprawnie, próbuje bowiem powstrzymać działanie technologii, której władza potencjalnie może być nieograniczona. Sztuczna inteligencja znalazłaby szybko rozwiązanie na ominięcie ograniczających ją praw.

Prawa robotów Asimova stały się pierwowzorem kolejnych prób regulacji problemu zasad moralnych, jakimi powinny się kierować urzędnicy wyposażone w sztuczną inteligencję. Od tego czasu powstało wiele koncepcji alternatywnych, stojących w opozycji do norm opracowanych przez tego amerykańskiego pisarza i profesora biochemii. Specyficzną ich odmianą jest koncepcja przyjaznej sztucznej inteligencji Eliezera Yudkowsky’ego, zakładająca, że można stworzyć roboty kierujące się moralnością, a tym samym przypisać im ludzkie cechy. Rolą ich twórców jest dbałość o to, aby relacje między ludźmi a robotami były harmonijne i aby uszanować fakt równorzędnego egzystowania jednych i drugich obok siebie¹². Można oczywiście tworzyć kolejne wytyczne, i takie działania podejmują już kolejne instytucje i państwa, ale przykład tych wcześniejszych prób wskazuje, że ich wielość nie poprawia poczucia bezpieczeństwa.

¹¹ I. Asimov, *Roboty i imperium*, tłum. P. Braiter-Ziemkiewicz, Rebis, Poznań 1994.

¹² *Etyka robotów*, https://pl.wikipedia.org/wiki/Etyka_robotów (21.12.2023).

Niewątpliwie AI może wspierać działania człowieka w różnorodnych obszarach jego funkcjonowania. Swoisty „technologiczny romantyzm” wywodzący się z kultury zachodniej wskazuje na sentymentalne odniesienie do technologii, wyrażające się w przekonaniu, że znakiem postępu jest rozwój technologiczny. Taki sposób myślenia tkwi swoimi korzeniami w przeświadczeniu, że technika i technologia wpływają bezpośrednio na przyspieszenie dynamizacji procesów ewolucyjnych ludzkości. Współcześnie w świecie, w którym coraz większą liczbę obszarów wspiera sztuczna inteligencja, trudno jest oderwać się od „ortodoksyjnego optymizmu” i poddać ją krytycznej analizie¹³. Pomimo rosnącej liczby przesłanek oceniających negatywnie wdrażanie tej nowej technologii¹⁴ zachowanie równowagi w rozwoju sztucznej inteligencji jest wciąż trudne. Możliwość jej kontrolowania i ustalania granic jej działalności jest związana bezpośrednio z umiejętnością dostrzegania jej zarówno pozytywnych, jak i negatywnych stron. Trudno jednak wierzyć w jej nieograniczone możliwości i jednocześnie wykazać się właściwym dystansem do jej roli w procesie wspierania rozwoju człowieka.

Ta rosnąca złożoność sprawia, że naukowcom trudniej niż kiedykolwiek zbadać, w jaki sposób postęp technologiczny zmienia życie na całym świecie zarówno w pozytywny, jak i negatywny sposób oraz jakie narzędzia społeczne, polityczne i prawne są potrzebne, aby wspierać kształtowanie rozwoju technologii w korzystnych kierunkach. Zadanie to może się wydawać trudne do wykonania w świetle szybkiego tempa zmian technologicznych i poczucia, że ich ciągły postęp jest nieunikniony. Mimo tego warto zdawać sobie sprawę zarówno z wyzwań, jak i znaczenia weryfikowania pełnego zakresu sposobów, w jakie technologia dokonuje transformacji społecznych, kulturowych, środowiskowych i politycznych. Wesprze to procesy planowania możliwych dróg kontrolowanego postępu sztucznej inteligencji i w następstwie kierowania nimi.

¹³ K. Łuszczek, *Wokół debaty „ed-tech”. Krytyka edukacyjnego wykorzystania nowych technologii*, „Dydaktyka Informatyki” 2018, nr 13, s. 70.

¹⁴ N. Selwyn, *Distrusting educational technology. Critical questions for changing time*, Routledge, New York 2013, s. 13.

‘Nie-masowa’ szkoła i syntetyczna edukacja

Współczesne społeczeństwo dzięki rozwojowi technologii w coraz szerszym zakresie egzystuje w kulturze masowej, a więc takiej, w której wytwory są produkowane i konsumowane przez szerokie rzesze ludności. Pojęcie kultury masowej odnosi się do norm i wzorców zachowania o bardzo rozległym zakresie zastosowania, ale również do mass mediów, gdyż dzięki masowym środkom komunikacji spełnione są w największym stopniu dwa podstawowe kryteria charakteryzujące kulturę masową: ilość i standaryzacja. Zdaniem Antoniny Kłoskowskiej kultura masowa jest produktem wtórnym rewolucji przemysłowej, a więc urbanizacji i industrializacji¹⁵. Mogła się ona rozwinąć min. dzięki technice umożliwiającej szybką dystrybucję treści i interakcji z odbiorcą, ale także dzięki upowszechnieniu oświaty na poziomie co najmniej elementarnym, a więc i pojawieniu się ‘masowego odbiorcy’.

Utworzenie szkoły masowej umożliwiło dostęp do edukacji wszystkim dzieciom. Bezpłatny system stworzył szansę na zdobycie wykształcenia niezależnie od pochodzenia czy statusu społecznego. Szkoła ogólnodostępna wpisana w kulturę masową zapewniła otwartość dla tych, którzy w innych warunkach nie mogliby z niej skorzystać, zmniejszyła również koszty edukacji. Niewątpliwie dużym minusem takiego systemu jest brak możliwości prowadzenia kształcenia zindywidualizowanego, co wyklucza rozwijanie indywidualnych zdolności czy zainteresowań. Wśród wad tego systemu wymienia się często: statyczność nauczanych treści, przymus edukacji, utrudniający osiąganie najlepszych możliwych rezultatów, narzucany autorytet nauczyciela, monotonię i powtarzalność, niemotywuujący system oceniania, system kar i nagród, pozbawiający uczniów stymulacji wewnętrznej, brak współpracy, nastawienie na sukces jednostki, promowanie potrzeby bycia lepszym od innych, brak wolności uczniów, brak kształtowania odpowiedzialności. Rosnąca liczba wad współczesnego systemu edukacji powoduje, że poszukiwane są metody jej reformy w kierunku, który daje jednostce szansę funkcjonowania nie tylko w obecnym świecie, ale również w przyszłości. Nowy system kształcenia powinien formować człowieka wolnego,

¹⁵ A. Kłoskowska, *Kultura masowa. Krytyka i obrona*, PWN, Warszawa 2005, s. 94.

odpowiedzialnego, kreatywnego, budującego relacje, współpracującego, potrafiącego motywować i uczyć się od siebie wzajemnie. Potrzebujemy edukacji dostosowanej do potrzeb i możliwości ucznia, zapewniającej optymalne warunki jego rozwoju.

W tej oczekiwanej modyfikacji systemów edukacyjnych znaczącą rolę może odegrać sztuczna inteligencja. Systemy komputerowe symulujące ludzkie myślenie mogą być szczególnie pomocne w indywidualizowaniu procesu nauczania czy indywidualizowaniu materiałów edukacyjnych, wspomagając tym samym dostosowanie procesu kształcenia do indywidualnych potrzeb i mocnych stron uczniów. Aplikacje AI mogą typować materiały dydaktyczne i podejścia dostosowane do potrzeb poszczególnych uczniów oraz przewidywać, rekomendować, a nawet podejmować decyzje dotyczące kolejnych etapów procesu uczenia się na podstawie analizy ich potrzeb i możliwości. Systemy sztucznej inteligencji mogą pomóc uczniom zdobywać wiedzę i umiejętności we własnym tempie, a jednocześnie dostarczać nauczycielom sugestii, w jaki sposób wspierać uczniów w procesie uczenia się¹⁶.

Wykorzystanie sztucznej inteligencji w edukacji nie jest niczym nowym. Jej historia sięga lat 60. XX wieku. Od tego czasu jest ona coraz szerzej integrowana z różnymi aspektami edukacji, od pomocy w nauczaniu arytmetyki i kształcenia umiejętności językowych, przez wypracowanie w latach 70. i 80. XX wieku systemów identyfikacji mocnych i słabych stron uczniów, po dostarczanie informacji zwrotnych dotyczących postępów w nauce poszczególnych uczniów. W latach 90. i w początkach XXI wieku opracowano pierwsze adaptacyjne systemy uczenia się, których celem jest indywidualizacja nauki przez dostosowanie treści i tempa nauczania do umiejętności i potrzeb konkretnych osób. W ostatnim czasie nastąpił szybki postęp w ulepszaniu platform edukacyjnych, które dopasowują się do potrzeb użytkowników i automatyzują projektowanie oraz produkcję materiałów edukacyjnych. Wraz z ewolucją sztucznej inteligencji będą doskonalone systemy wsparcia kształcenia, a tym samym będzie ona odgrywała coraz ważniejszą rolę w kształtowaniu edukacji przyszłości. Będzie indywidualizowała nie tylko proces nauczania, ale także sposoby przygotowania niezbędnych materiałów

¹⁶ Por. S. Jaskuła, *Sztuczna inteligencja*, w: *Rzeczywistość hybrydalna. Perspektywa wychowawcza*, red. S. Jaskuła, Księgarnia Akademicka, Kraków 2023, s. 95–112.

dydaktycznych. W ten sposób mogą powstawać indywidualne podręczniki, uwzględniające potrzebę elastycznych zmian treści i metod.

Sztuczna inteligencja może być wykorzystywana w procesach kształcenia również do:

- a) tworzenia wirtualnych tutorów dla każdego ucznia,
- b) analizy danych dotyczących interakcji w edukacji,
- c) automatyzacji podstawowych prac administracyjnych podejmowanych przez nauczycieli,
- d) ułatwiania dostępności treści edukacyjnych,
- e) wspierania procesów preorientacji zawodowej,
- f) wspierania edukacji inkluzyjnej itp.

Przejmowanie procesów kształcenia przez sztuczną inteligencję wraz z jej rozwojem będzie rodzić coraz więcej pytań o to, czy może ona oddziaływać na inteligencję ludzi, a przede wszystkim: czy powinna. Czy ułatwianie i uprzyjemnianie nauki, która dotąd była wyzwaniem, nie doprowadzi do tego, że jednostki będą gorzej przygotowane do radzenia sobie w realnym życiu? Jakie długofalowe skutki wykorzystania AI w procesach kształcenia zaczniemy dostrzegać dopiero za kilkanaście lat? I wreszcie czy będziemy mogli kontrolować to, czego faktycznie będzie uczyć sztuczna inteligencja, a w konsekwencji: jakiego człowieka ukształtuje?¹⁷

Sztuczna inteligencja wykorzystywana w procesach kształcenia już dzisiaj koncentruje się na budowaniu zasobów danych, na podstawie których uczeń tworzy informacje i buduje swoją wiedzę. Czy rozwija jednak swoją samoświadomość, tożsamość i nabywa kompetencje ważne dla własnego rozwoju? Nikt nie pyta, czemu mogą służyć nowe umiejętności, wiedza czy informacje? W jaki sposób są one wykorzystywane w całokształcie osobowego dojrzewania jednostki? Obecny proces kształcenia, tym bardziej z wykorzystaniem sztucznej inteligencji, koncentruje się na analizie postępu jego cech formalnych, poddawanych następnie różnorodnym technikom testowania. Sztuczna inteligencja wykorzystywana w procesie kształcenia dokonuje weryfikacji poziomu nabycia umiejętności instrumentalnych, marginalizując problem wychowawczy, uciekając od ich wymiaru etycznego oraz pytania,

¹⁷ Por. S. Jaskuła, *Sztuczna inteligencja w edukacji we współczesnej rzeczywistości hybrydalnej*, „Perspektywy Kultury” 2023, nr 42 (3), s. 13–26.

w jakim stopniu są one przydatne integralnie rozumianemu człowiekowi i ludzkości. Tym samym współczesna edukacja, a jeszcze silniej ta, w której stosujemy sztuczną inteligencję, coraz szybciej zostaje przystosowana do tego, o czym pisał Erich Fromm, tj. do „posiadania”, względnie do sprawnego funkcjonowania, ale nie „bycia”. W książce *Być czy mieć?* autor wyróżnia te dwa sposoby egzystencji: modus bycia i posiadania. Krytykuje on współczesne społeczeństwo w głównej mierze dlatego, że jest ono nastawione na modus posiadania. Nawiązuje do tego problemu w kolejnej monografii, *O sztuce istnienia*: „W świecie, w którym dominującą cechą egzystencji jest modus posiadania, naczelne motto brzmi: ‘Jestem tym, co mam’. Po dokonanej transformacji przybierze ono postać: ‘Jestem tym, co robię’ (...) lub po prostu: ‘Jestem tym, kim jestem’”¹⁸. Modus posiadania jest egocentrycznym sposobem życia, partykularnie nastawionym do świata i do siebie, jest to egoistyczna postawa, która odzwierciedla całokształt myślenia, odczuwania i działania, nastawiona na adaptację i wypełnianie wymogów przypisanych ról społecznych. Z kolei bycie związane jest z przeżyciem, pasją życia i wolnością jednostki. Taki człowiek aktywnie istnieje, jest twórczy, dokonuje się w nim rozwój definiowany w sposób znacznie bardziej złożony, czasami nawet wbrew funkcjonalnym wymogom środowiska, zdolny do rozpoznawania nie tylko potrzeb, ale i wartości, rozpoznający transgresje pozytywne i negatywne: „Być wymaga rezygnacji z egocentryzmu i egoizmu, mówiąc zaś słowami mistyków – wymaga ‘pustki’ i ‘ubóstwa’”¹⁹. Modus „mieć” nakierowany jest przede wszystkim na wartości materialne – modus „być” – na wartości humanistyczne, np. na miłość, godność czy altruizm, a więc na wartości autoteliczne²⁰.

Edukacja wykorzystująca sztuczną inteligencję zagrożona jest ukierunkowaniem na modus „mieć”, a tym samym na zawłaszczanie, nabywanie i działanie samolubne, bez dzielenia się czy współpracy. Jednostka egzystująca w takim modelu podporządkowuje proces uczenia się posiadaniu jak największej ilości informacji czy umiejętności, ale tylko dla siebie i z przeznaczeniem na swój użytek²¹. Inną cechą tego modułu jest jego utylitarny,

¹⁸ E. Fromm, *O sztuce istnienia*, Vis-à-Vis Etiuda, Warszawa 2023, s. 130.

¹⁹ Tamże, s. 149.

²⁰ Por. tamże, s. 199–203.

²¹ Por. E. Fromm, *Mieć czy być?*, Rebis, Poznań 2022, s. 66.

rzeczowy charakter, który rezygnuje z atrybutu podmiotowości człowieka. Celem procesu edukacji podporządkowanej tak konstruowanej strategii przestaje być samodoskonalenie, staje się nim za to zdobywanie umiejętności pozwalających posiadać jeszcze więcej. Nie rozwijanie samodzielności, lecz wspieranie się sztuczną inteligencją, która przecież ma dostęp do wszystkich danych, by na ich podstawie rozwiązywać dowolne problemy. Zwalnia to uczestników edukacji z odpowiedzialności za własne uczenie się i samorozwój. To sztuczna inteligencja nadzoruje i kontroluje te procesy, wskazując na braki i dostarczając kolejne treści. Dodatkową konsekwencją tak pojętego nauczania jest niemal zupełna ucieczka od problemów wychowawczych, których identyfikacja wymaga odejścia od tak radykalnie zredukowanej aksjologii i duchowości człowieka, od fascynacji erą posthumanizmu.

Pytaniami otwartymi pozostają: Czy o taką edukację nam chodzi? Czy tak ma wyglądać człowiek przyszłości? Niesamodzielny samotnik nastawiony na egoistyczne posiadanie? Być może obraz ten jest przesadzony, ale pokazuje, dokąd może zmierzać edukacja, jeśli pozwolimy sztucznej inteligencji na przejęcie procesów kształcenia i wychowania, jeśli zmarginalizujemy człowieka zarówno w procesie nauczania, jak i uczenia się.

Obecne zacieranie różnic między „cyfrowymi tubylcami” a „cyfrowymi emigrantami” nie zwalnia nas z dostrzegania szkody rozwojowej, jaką może przynieść sztuczna inteligencja. Technologia nie może wypełniać całkowicie środowiska wychowawczego i być głównym czynnikiem kształtującym szczególnie młodego człowieka. Tak jak „robotyka społeczna”, wykorzystująca złudzenia, jakim ulega człowiek w konfrontacji z maszyną, nie może być podstawą w kształtowaniu relacji interpersonalnych i konstytuujących relacje człowieka z maszynami. Efekt „inteligencji emocjonalnej”, czyli wyposażenia maszyn w umiejętność rozpoznawania i wyrażania emocji, będzie niestety pogłębiał relacje człowieka z maszyną i niwelował potrzebę kontaktu z drugim człowiekiem. Warto jednak pamiętać, że emocje to jeszcze nie wartości, że ich charakter może być bardzo odmienny. Mogą bazować zarówno na doraźnie odczuwalnym stresie, lęku i zagubieniu, jak i na uczuciach wyższego rzędu, odwołujących się do wspomnianych wartości autotelicznych. Behawioralne korelaty emocji, które rozpoznać i naśladować może komputer, nie wystarczą do zdefiniowania ich treści, do budowania motywacji ich przekraczania.

Zakończenie

Stoimy na rozdrożu: jedna z dróg prowadzi do całkowicie zmechanizowanego społeczeństwa, w którym człowiek jest bezbronnym trybikiem wielkiej maszyny i równie dobrze może doprowadzić do zagłady (...); druga zaś z dróg prowadzi do odrodzenia humanizmu i nadziei – do społeczeństwa, które stawia technikę w służbie dobru człowieka²².

Cytat ten pokazuje, że współcześnie istnieje coraz poważniejszy problem egzystencji człowieka w zdehumanizowanym świecie, ale i tu – Fromm zawsze jednak był optymistą – widzi dla ludzkości... nadzieję.

Wychowanie do człowieczeństwa jest nam coraz bardziej potrzebne, podobnie jak misja ochrony człowieczeństwa dla kolejnych pokoleń, dla zachowania świata. Sztuczna inteligencja w tym zakresie niewiele nam pomoże, a wręcz przeciwnie. Nie oznacza to jednak jej nieużyteczności i braku możliwości zastosowania w procesach edukacji. Zapewne szerzej w procesach kształcenia niż wychowania. W tym drugim procesie poważnym problemem może być ustalenie, kto wyznacza cele wychowawcze, kto jest autorytetem i wzorem do naśladowania. Jak bez udziału człowieka możliwe jest zrozumienie ludzkich wartości? Czy wreszcie: Jak możliwe są uspołecznienie i socjalizacja niezaprogramowane przez adaptacyjne technologie kultur organizacyjnych, które opanowały już wszystkie dziedziny życia? Jak możliwe jest budowanie relacji międzyludzkich przez technologie?

Włączając sztuczną inteligencję do edukacji, musimy zdawać sobie sprawę z różnych zagrożeń, i w tym przypadku prosta zależność przewagi aspektów pozytywnych nad negatywnymi nie pomoże w jednoznacznej odpowiedzi na pytanie, czy jest to wystarczający argument za podjęciem ryzyka zastosowania technologii w kształceniu lub wychowaniu. Niezależnie od tych rozważań nie ulega wątpliwości, że jako ludzkość nie możemy przeoczyć momentu, w którym powstaną maszyny o inteligencji przewyższającej intelekt człowieka, bo wtedy to nie my, ludzie, będziemy decydować o ich faktycznej roli w edukacji.

²² E. Fromm, *Rewolucja nadziei. W stronę uczonej technologii*, Vis-à-Vis Etiuda, Kraków 2013.

Bibliografia

- Andreu A., Ravindran J., *Ludzie nie będą zdolni do kontrolowania superinteligentnej AI. To wnioski z badania*, 6.10.2022, <https://businessinsider.com.pl/technologie/nowe-technologie/naukowcy-sztuczna-inteligencja-bedzie-nie-do-opanowania/qewm60w> (21.12.2023).
- Asimov I., *Roboty i imperium*, tłum. P. Braiter-Ziemkiewicz, Rebis, Poznań 1994.
- Asimov I., *Zabawa w berka*, w: tegoż, *Ja, robot*, tłum. Z. A. Królicki, Rebis, Poznań 1980.
- Bell D., *The Measurement of Knowledge and Technology*, w: *Indicators of Social Change*, ed. E. Sheldon, W. Moore, Russell Sage Foundation, New York 1968.
- Etyka robotów*, https://pl.wikipedia.org/wiki/Etyka_robotów (21.12.2023).
- Fromm E., *Mieć czy być?*, Rebis, Poznań 2022.
- Fromm E., *O sztuce istnienia*, Vis-à-Vis/Etiuda, Warszawa 2023.
- Fromm E., *Revolucja nadziei. W stronę uczłowieczonej technologii*, Vis-à-Vis/Etiuda, Kraków 2013.
- Fromm E., *Zdrowe społeczeństwo*, tłum. A. Tanalska-Dulęba, Vis-à-Vis/Etiuda, Warszawa 1996.
- Gunning D., Aha D. W., *DARPA's Explainable Artificial Intelligence Program*, „AI Magazine” 2019, 40(2), s. 44–58.
- Jaskuła S., *Jak sztuczna inteligencja zmienia świat*, w: *Świat w czasie przełomu. Religijne mosty nadziei*, red. P. Walewski, Bernardinum, Pelplin 2023, s. 53–69.
- Jaskuła S., *Sztuczna inteligencja*, w: *Rzeczywistość hybrydalna. Perspektywa wychowawcza*, red. S. Jaskuła, Księgarnia Akademicka, Kraków 2023, s. 95–112.
- Jaskuła S., *Sztuczna inteligencja w edukacji we współczesnej rzeczywistości hybrydalnej*, „Perspektywy Kultury” 2023, nr 42(3), s. 13–26.
- Kłосkowska A., *Kultura masowa. Krytyka i obrona*, PWN, Warszawa 2005.
- Łuszczek K., *Wokół debaty „ed-tech”. Krytyka edukacyjnego wykorzystania nowych technologii*, „Dydaktyka Informatyki” 2018, nr 13, s. 67–75.
- Selwyn N., *Distrusting educational technology: Critical questions for changing time*, Routledge, New York 2013.
- Zwoliński A., *Cyberpułapka. Jak internet, sztuczna inteligencja i wirtualna rzeczywistość skutecznie niszczą cywilizację*, Fronda, Warszawa 2023.

Artificial teacher – true educational consequences

Summary

Artificial intelligence is increasingly seen as a central element of digital social transformation and the digitization of many sectors. Of all the areas of life that will have a fundamental impact, education may be the most important. Artificial intelligence has the potential to solve some of the biggest challenges in teaching today, innovate teaching and learning practices, and ultimately accelerate progress in these areas. Currently developing artificial intelligence can significantly change current educational models, which, despite constant but insignificant evolution, still remain static, referring to old patterns that are no longer current. In a special way, AI can support the transformation of the mass education system towards its individualization, but not only that. On the other hand, education can play a significant role in eliminating the threats that already accompany the development of artificial intelligence.

Although there are many advantages resulting from the implementation of artificial intelligence in educational processes, its use in educational processes raises significant moral and social concerns at this stage. Concerns about depersonalization and ethical considerations should be at the forefront of discussions about the use of artificial intelligence in education, highlighting the need to carefully consider the scope of its implementation.

Keywords: artificial intelligence, artificial teacher, educational processes

Sztuczna inteligencja jako substytut „religii” współczesnego człowieka. Wyniki badań socjologicznych

Wprowadzenie

Współczesny świat, naznaczony dynamicznym rozwojem technologii sztucznej inteligencji, implikując wiele zmian społecznych, stanowi nowe wyzwanie dla tradycyjnych instytucji, w tym m.in. dla Kościoła katolickiego. Postępujący proces laicyzacji społeczeństwa polskiego oraz ewolucja społeczno-kulturowa zdają się kształtować nową rzeczywistość, która wchodzi w dialog z doktryną religijną³. Zainteresowanie religijną sferą polskiego życia społecznego w kontekście wykorzystania technologii sztucznej inteligencji (dalej: SI) implikowane jest charakterem analizowanego społeczeństwa, które – mimo postępującej sekularyzacji⁴ – jest utożsamiane z wyznaniem rzymskokatolickim. Jednocześnie intensywny rozwój i coraz częstsze wykorzystanie technologii AI zmienia jej

¹ Dr hab. Małgorzata Gruchola, prof. KUL, Wydział Nauk Społecznych, Katolicki Uniwersytet Lubelski Jana Pawła II; Instytut Badań Edukacyjnych w Warszawie.

² Dr Robert Zieliński, Instytut Badań Edukacyjnych w Warszawie.

³ M. Przybyś, *Nie znam człowieka nawróconego dzięki sztucznej inteligencji. Ale może to kwestia czasu?*, <https://pl.aleteia.org/2023/01/21/sztuczna-inteligencja-w-kosciele/> (2.01.2024).

⁴ Główny Urząd Statystyczny, *Wyznania religijne w Polsce w latach 2019-2021. Analizy statystyczne*, GUS, Warszawa 2022, s. 67. Kościół Rzymskokatolicki w Polsce: 32190661. Spadek wiernych w % ogółu ludności z 95,5% w 2000 r. do 92,2% w 2021 r.

status i postrzeganie. Coraz częściej z wytworu człowieka: narzędzia, przedmiotu zmienia się w „podmiot” kultu religijnego. Z tej racji badania empiryczne ukierunkowane są na weryfikację wpływu oddziaływania technologii sztucznej inteligencji na praktyki religijne badanych Polaków.

Celem tekstu jest wykazanie – przez analizę wyników badań empirycznych – że sztuczna inteligencja jest substytutem nowej religii współczesnego człowieka, rozumianej w ujęciu socjologicznym. Religia jest definiowana jako relacja człowieka do różnie określanej świętości (*sacrum*), a także sfery boskiej, manifestująca się w wymiarze doktrynalnym (wiara), w czynnościach religijnych (kult, obrzędy, rytuały), w sferze społeczno-organizacyjnej (m.in. Kościół, wspólnota religijna) i w religijności indywidualnej (m.in. w mistyce)⁵. Socjologia, w odróżnieniu od teologii, koncentruje się na społecznym aspekcie zjawiska religijnego i bada nie tylko przekonania, lecz również ich zobiektywizowane formy wyrazu, mianowicie kult i organizację⁶.

Socjologiczne definicje religii możemy podzielić na przedmiotowe i funkcjonalne. Definicje przedmiotowe koncentrują się na przedmiocie czci religijnej i wiary. Stanowi go – w zależności od poglądów autora definicji – Bóg (Max Scheler, Søren Kierkegaard), bóstwa (Wilhelm Schmidt, Edward Tylor), to, co nadzmysłowe (Franciszek Mantey), nadprzyrodzone (Cornelius Peter Tiele, Nathan Soderblom), rzeczywistość nieempiryczna (Władysław Piwowski), nieskończoność (Franz Adler, Friedrich Schleiermacher) oraz świętość (Rudolf Otto)⁷. Definicje funkcjonalne natomiast ujmują religię ze względu na pełnione przez nią funkcje⁸. Religia nadaje życiu sens, pełni funkcję więziotwórczą i integracyjną, wychowawczą i edukacyjną, psychologiczną i egzystencjalną (Émile Durkheim), zespala społeczeństwo wokół określonych wartości, sprawuje kontrolę społeczną, obiecuje wsparcie psychiczne (Norman Goodman)⁹.

⁵ *Religia*, w: *Encyklopedia PWN*, <http://encyklopedia.pwn.pl/haslo/religia;3966983.html> (28.12.2023).

⁶ W. Pawluczuk, *Religia*, w: Z. Bokszański (red.), *Encyklopedia socjologii*, red. Z. Bokszański, t. 3, Oficyna Naukowa, Warszawa 2000, s. 288.

⁷ Tamże, s. 289.

⁸ Zob. N. Goodman, *Wstęp do socjologii*, tłum. J. Polak, J. Raszkowski, U. Zielińska, Zysk i S-ka, Poznań 2009, s. 217.

⁹ Zob. M. Gruchola, *Konsumpcjonizm – nową religią współczesnego człowieka?*, „Roczniki Kulturoznawcze” 2017, t. 8, nr 3, s. 39–62.

Problem badawczy koncentruje się wokół terminologii. Jak słusznie zauważa Krzysztof Różanowski, pojęcie sztucznej inteligencji, „mimo powszechności używania tego terminu, nie jest łatwe do zdefiniowania”¹⁰. Aktualnie są pewne kłopoty z określeniem, „co kryje się pod ogólnym pojęciem «sztucznej inteligencji»”¹¹. Stąd „próba jednoznacznego jej zdefiniowania jest to – paradoksalnie – działanie «sztuczne» [...]”¹².

Sformułowanie jednoznacznej i niebudzącej zastrzeżeń definicji sztucznej inteligencji nie jest łatwe, czego dowodzi np. dyskusja wokół charakterystyki AI zawartej w unijnym projekcie rozporządzenia AI Act. Sytuację komplikuje również to, że zagadnienia poruszane w ramach etyki AI nierzadko obejmują dziedziny, których AI jest tylko komponentem, jak np. robotyka czy internet rzeczy¹³.

Klasycznie literatura przedmiotu – za Johnem R. Searle’em – wydziela dziedzinę sztucznej inteligencji na dwie sekcje: silną i słabą. Pierwsza – ogólna AI – ujmuje sztuczną inteligencję jako „[h]ipotetyczne, sztuczne systemy inteligentne o charakterze kompleksowym, lub przynajmniej wielozadaniowym, które w autonomiczny sposób funkcjonują w środowisku, wykazując się inteligencją”. Z kolei słaba, wąska AI obejmuje: „programy wykonujące konkretne zadania lub rozwiązujące konkretne problemy, które w przypadku, gdy realizowane czy rozwiązywane są przez ludzi, uznaje się za wymagające wykazania się inteligencją”¹⁴.

Na wstępie procesu badawczego przyjęto hipotezę, że sztuczna inteligencja może stać się substytutem religii współczesnego człowieka.

¹⁰ K. Różanowski, *Sztuczna inteligencja: rozwój, szanse i zagrożenia*, „Zeszyty Naukowe” 2007, nr 2, s. 110.

¹¹ F. Kurp, *Sztuczna inteligencja od podstaw*, Helion, Gliwice 2023, s. 11.

¹² M. Nowakowski, *O moralnej odpowiedzialności HAL-a 90001, czyli etyka sztucznej inteligencji w praktyce. Czy potrzebujemy definicji sztucznej inteligencji?*, „Prawo Mediów Elektronicznych” 2022, nr 1, s. 7.

¹³ M. Chojnowski, *Etyka sztucznej inteligencji*, Centrum Etyki Technologii Instytutu Humanities, Warszawa 2022, s. 8.

¹⁴ J. R. Searle, *Minds, brains, and programs*, „The Behavioral and Brain Sciences” 1980, no. 3, s. 417.

Metoda i materiał badawczy

W artykule zostaną zaprezentowane wybrane wyniki ogólnopolskich badań, które zostały przeprowadzone przez Instytut Badań Edukacyjnych we wrześniu 2023 r. na próbie 1052 osób w wieku 15–65 lat¹⁵. Uwzględniono dwa bloki pytań: pytania dotyczące postaw Polek i Polaków wobec religijności i modlitwy oraz pytania dotyczące ich postaw wobec sztucznej inteligencji. W badaniach zastosowano metodę ilościową, technikę kwestionariusza ankiety. Narzędzie badawcze zostało zbudowane przez zespół badaczy na potrzeby analizy dotyczącej poruszanej w artykule problematyki. Składało się z: 15 pytań jednokrotnego wyboru bazujących na krótkiej kafeterii (przy zastosowaniu zazwyczaj skali Likerta), 2 pytań wielokrotnych odpowiedzi (6-stopniowa kafeteria) i 1 pytania składającego się z 26 itemów ocenianych na skali 1–5. Ponadto zastosowano metryczkę zawierającą informację na temat cech demograficzno-społecznych. Kolportaż narzędzi badawczych odbył się przy zastosowaniu panelu CAWI (Computer Assisted Web Interview). W dobrze próby badawczej zastosowano metodę losowo-kwotową ze względu na zwiększenie prawdopodobieństwa w celu odwzorowania cech całej populacji.

Jako zmienną niezależną przyjęto praktyki z programami opartymi na technologii sztucznej inteligencji (praktyki z SI), praktyki religijne i praktyki codziennej modlitwy. Zmienne te poddano analizie ze zmiennymi z grupy cech demograficznych. Należy podkreślić, iż w większości przypadków analiz statystycznych w korelacji ze zmiennymi zależnymi stwierdzono istotność asymptotyczną (dwustronną) na poziomie 0,01 lub 0,05 w testach opartych na χ^2 Pearsona.

Wyniki badań

W badaniu wzięło udział 52,1% kobiet i 47,9% mężczyzn, wśród których 2,6% zadeklarowało wykształcenie podstawowe/gimnazjalne, 7,3% – zasadnicze, 34,3% – średnie, 12,8% – pomaturalne/policealne, 10,7% – wyższe

¹⁵ M. Gruchola, R. Zieliński, M. Zajic, J. Koper, M. Popielewicz-Durakiewicz, *Badanie postaw Polek i Polaków wobec AI*, Zakład Edukacji Medialnej i Sztucznej Inteligencji IBE, Warszawa 2023.

licencjackie i 32,2% – wyższe magisterskie i równorzędne. Spośród respondentów 17,8% zamieszkuje wieś, 13,6% – małe miasto (do 20 tys. mieszkańców), 15,8% – średnie miasto (od 20 do 99 tys. mieszkańców), 43,7% – duże miasto (od 100 do 500 tys. mieszkańców), a 9,1% – wielkie miasto (powyżej 500 tys. mieszkańców). Poniżej przedstawiono interpretację badań z uwzględnieniem zmiennych niezależnych praktyki korzystania z SI, a także praktyki modlitw codziennych, które przybliżają spektrum omawianej problematyki. Jako zmienne zależne wzięto pod uwagę cechy demograficzne i cechy społeczne oscylujące wokół religijności.

Tabela 1. Cechy demograficzne a praktyki z AI (dane w %)

Cechy demograficzne		Praktyki korzystania z SI						Ogółem
		Codziennie	Kilka razy w tygodniu	Raz w tygodniu	Jeden, dwa razy w miesiącu	Rzadziej niż jeden raz w miesiącu	Nigdy nie korzystałem	
Płeć	Kobieta	16,0	10,4	5,6	8,2	15,7	44,2	100,0
	Mężczyzna	10,8	16,1	5,9	12,1	16,8	38,4	100,0
Wiek	15–24 lata	20,8	19,8	7,0	18,5	16,5	17,5	100,0
	25–34 lata	11,6	16,3	6,1	11,3	18,2	36,5	100,0
	35–44 lata	20,7	13,0	6,2	7,6	12,4	40,2	100,0
	45–54 lata	13,4	11,0	2,3	11,5	22,5	39,3	100,0
	55–65 lat	7,7	9,0	6,3	5,6	13,5	57,9	100,0
Wykształcenie	Podstawowe / gimnazjalne	9,7	29,4	3,5	11,0	14,9	31,6	100,0
	Zasadnicze	13,7	16,4	1,6	5,4	9,7	53,2	100,0
	Średnie	12,9	12,5	4,5	11,8	17,1	41,3	100,0
	Pomaturalne / policealne	19,0	9,7	9,8	10,4	9,3	41,8	100,0
	Wyższe licencjackie	7,8	17,5	10,8	13,1	14,3	36,5	100,0
	Wyższe magisterskie	14,3	11,6	4,8	8,0	20,4	41,0	100,0
Miejscowość zamieszkania	Wieś	12,7	11,0	5,0	11,9	14,4	44,9	100,0
	Małe miasto (do 20 tys. mieszkańców)	10,7	13,7	10,0	10,8	15,8	39,0	100,0
	Średnie miasto (od 20 do 99 tys. mieszkańców)	18,4	11,1	5,5	8,6	18,6	37,8	100,0
	Duże miasto (od 100 do 500 tys. mieszkańców)	13,5	13,7	4,6	8,2	16,8	43,2	100,0
	Wielkie miasto (powyżej 500 tys.)	11,1	17,1	6,1	16,7	13,7	35,3	100,0

Źródło: opracowanie własne

W tabeli 1 przedstawiono analizy statystyczne uwzględniające podstawowe cechy demograficzne badanych (wiek, płeć, wykształcenie, miejscowość). Szczegółowe analizy statystyczne wskazują na istotny związek zmiennych z grupy cech demograficznych (istotność asymptotyczna na poziomie 0,01), oprócz miejscowości zamieszkania ($\chi^2 = 23,096$, $df = 20$, $p = 0,284$). Istnieje różnica w częstotliwości korzystania z AI między płciami – kobiety częściej niż mężczyźni deklarują brak doświadczenia ze sztuczną inteligencją. Analiza wskazuje, że wśród respondentek prawie połowa (44,2%) nigdy nie korzystała z AI, podczas gdy 16,0% deklaruje korzystanie codzienne. Z kolei wśród płci męskiej 38,4% badanych nigdy nie miało do czynienia z AI, a 16,1% korzysta z niej kilka razy w tygodniu. Wartości procentowe w poszczególnych kategoriach ujawniają, że do regularnego korzystania z technologii AI bardziej skłonni są mężczyźni niż kobiety.

Analiza danych dotycząca związku między wiekiem respondentów a korzystaniem ze sztucznej inteligencji wykazuje istotne zależności. Najmłodsza grupa wiekowa, czyli osoby w przedziale 15–24 lata, wykazuje się największym zainteresowaniem sztuczną inteligencją. W tej grupie 20,8% respondentów korzysta z AI codziennie, a 19,8% kilka razy w tygodniu. Warto jednak zauważyć, że 17,5% osób w tym przedziale wiekowym nie miało do czynienia z AI nigdy. W grupie wiekowej 25–34 lata można zaobserwować zróżnicowane podejście. Mimo że 36,5% respondentów nigdy nie korzystało z AI, to 18,2% deklaruje korzystanie z tej technologii kilka razy w tygodniu. W przypadku osób w wieku 35–44 lata aż 40,2% respondentów zgłasza, że nigdy nie korzystało z AI, co stanowi najwyższy odsetek w tej kategorii. W grupie wiekowej 45–54 lata większość respondentów (39,3%) również deklaruje brak doświadczenia z AI, ale znaczący odsetek badanych (22,5%) korzysta z tej technologii rzadziej niż raz w miesiącu. Najstarsza grupa wiekowa, czyli osoby w wieku 55–65 lat, prezentuje się jako najmniej aktywna w korzystaniu z AI. Tu ponad połowa (57,9%) respondentów nigdy nie korzystała z tej technologii. Wartości procentowe zainteresowania i aktywności w korzystaniu ze sztucznej inteligencji maleją wraz z wyższym wiekiem respondentów. Młodsze pokolenia są bardziej otwarte na regularne korzystanie z AI, podczas gdy starsze – mniej w tym obszarze aktywne. Grupa wiekowa 35–44 lata stanowi pewien wyjątek; tu brak doświadczenia z AI deklaruje duża część respondentów, co może wynikać z takich

czynników, jak specyfika zawodowa czy stopień zaznajomienia z technologią w tej grupie wiekowej podyktowany brakiem zainteresowania nowościami opartymi na technologiach cyfrowych.

Weryfikując dane dotyczące korzystania z technologii AI w zależności od poziomu wykształcenia respondentów, można zauważyć kilka istotnych tendencji. Osoby z wykształceniem podstawowym lub gimnazjalnym rzadziej korzystają z AI, zaledwie 9,7% z nich używa ich codziennie, a 31,6% nigdy nie miało z nimi do czynienia. W przypadku osób z wykształceniem zasadniczym 53,2% nie korzystało z AI nigdy, co sugeruje słabszą znajomość technologii w tej grupie. Osoby z wyższym wykształceniem, zwłaszcza licencjackim i magisterskim, częściej korzystają z analizowanej technologii. Wśród osób z wykształceniem wyższym licencjackim 17,5% korzysta z AI kilka razy w tygodniu, podczas gdy dla osób z wykształceniem magisterskim odsetek ten wynosi 11,6%. Należy zauważyć, że 14,3% respondentów z wykształceniem magisterskim używa AI codziennie. Nasuwa się wniosek, że im wyższy poziom wykształcenia, tym większa aktywność w korzystaniu z technologii AI.

Wśród mieszkańców wsi zaobserwowano niski wskaźnik korzystania z AI, aż 44,9% badanych nigdy nie miało z nimi do czynienia. Podobna tendencja jest widoczna w małych miastach, gdzie z AI nie korzysta 39% respondentów. Warto zauważyć, że w mniejszych miejscowościach obserwuje się niższy udział osób korzystających z technologii na co dzień. Mieszkańcy średnich miast wykazują natomiast większą zmienność w praktykach z AI, z odsetkiem osób korzystających z technologii codziennie wynoszącym 18,4%. Duże i wielkie miasta prezentują wyższy odsetek codziennego użytkowania AI, odpowiednio 13,5% i 11,1%. Dodatkowo częstotliwość korzystania ze sztucznej inteligencji kilka razy w tygodniu mieszkańcy dużych miast wykazują w większym stopniu niż mieszkańcy mniejszych miejscowości. Należy podkreślić, iż praktyki te mogą odzwierciedlać zarówno dostępność infrastruktury technologicznej, jak i świadomość cyfrową społeczności. Nasuwa się wniosek, że obszary miejskie, zwłaszcza te większe, wydają się sprzyjać większej aktywności w korzystaniu z AI, co może wynikać z dostępności usług oraz szerokiego zakresu możliwości, jakie oferują cyfrowe technologie, a w szczególności internet, charakteryzujący się bardzo dużym przesyłem megabajtów (światłowód).

Bardzo znaczące wskaźniki, które mają istotny wpływ na praktyki z aplikacjami bazującymi na sztucznej inteligencji, należą do grupy cech społecznych. W tym wypadku uwzględnia się elementy religijne, czyli auto-identyfikację religijną, praktyki religijne i praktyki dotyczące codziennej modlitwy. Analiza danych pozwala na wyciągnięcie kilku istotnych wniosków. Osoby deklarujące się jako wierzące wydają się mniej skłonne do korzystania z aplikacji AI. Ponad 82% z nich nigdy nie sięga po tego typu technologie, co wskazuje na pewne zahamowania w akceptacji nowoczesnych rozwiązań technologicznych w tej grupie. Z kolei im mniejsze zaangażowanie religijne, tym wyższa jest częstotliwość korzystania z aplikacji sztucznej inteligencji. Osoby, które deklarują, że są niewierzące, prezentują najwyższe wskaźniki korzystania z tej technologii (codziennie: 24,0%), co sugeruje większą otwartość na nowoczesne rozwiązania w tej grupie. Niezdecydowani i poszukujący oscylują w środku skali, prezentując umiarkowane poziomy korzystania z aplikacji AI. W tej grupie najczęściej korzysta się z technologii kilka razy w tygodniu (odpowiednio 26,6% i 10,1%). Przeprowadzone analizy pozwalają sformułować wniosek, że większość respondentów korzysta z aplikacji sztucznej inteligencji z umiarkowaną częstotliwością, przy czym najczęściej jest to kilka razy w tygodniu. Warto jednak podkreślić, że analizowane dane ukazują pewne zróżnicowanie w zależności od cech religijnych respondentów, sugerując, że poziom zaangażowania religijnego może wpływać na akceptację programów bazujących na technologii AI i korzystanie z nich.

Tabela 2. Cechy religijne a praktyki z AI (dane w %)

Cechy społeczne oscylujące wokół religijności		Praktyki korzystania z AI						Ogółem
		Co-dziennie	Kilka razy w tygodniu	Raz w tygodniu	Jeden, dwa razy w miesiącu	Rzadziej niż jeden raz w miesiącu	Nigdy nie korzystałem	
Autoidentyfikacja religijna	Głęboko wierzący	7,2	14,4	4,5	9,0	27,9	36,9	100,0
	Wierzący	11,1	12,4	6,5	9,9	14,7	45,4	100,0
	Niezdecydowany	13,7	26,6	4,8	7,3	12,1	35,5	100,0
	Poszukujący	15,2	10,1	8,9	8,9	12,7	44,3	100,0
	Niewierzący	24,0	6,6	3,3	13,7	18,6	33,9	100,0

Cechy społeczne optymalizujące wokół religijności		Praktyki korzystania z AI						Ogółem
		Co- dziennie	Kilka razy w tygo- dniu	Raz w tygo- dniu	Jeden, dwa razy w mie- siącu	Rzadziej niż jeden raz w miesiącu	Nigdy nie korzy- stałem	
Praktyki religijne	Kilka razy w tygodniu	3,6	3,6	5,5	9,1	34,5	43,6	100,0
	Raz w tygodniu	12,6	13,8	5,7	8,9	13,8	45,1	100,0
	Jeden, dwa razy w miesiącu	11,4	25,0	9,1	10,2	21,6	22,7	100,0
	Jeden raz, kilka razy w roku	12,1	13,7	4,8	12,9	12,1	44,4	100,0
	Rzadziej niż raz w roku	12,6	9,2	7,6	11,8	21,8	37,0	100,0
	W ogóle w nich nie uczestniczę	18,3	12,0	4,7	8,0	14,3	42,7	100,0
Praktyki codziennej modlitwy	Kilka razy dziennie	12,7	10,2	3,4	9,3	23,7	40,7	100,0
	Raz dziennie	14,2	16,8	7,6	4,1	13,2	44,2	100,0
	Kilka razy w tygodniu	12,5	20,8	5,8	10,8	22,5	27,5	100,0
	Raz w tygodniu	7,5	9,4	22,6	3,8	11,3	45,3	100,0
	Jeden, dwa razy w miesiącu	10,7	16,7	2,4	20,2	22,6	27,4	100,0
	Rzadziej niż jeden raz w miesiącu	9,7	9,7	3,9	11,7	14,6	50,5	100,0
	Nie praktykuję modlitwy	16,2	10,1	4,2	11,4	13,3	44,8	100,0

Źródło: opracowanie własne

Dokonując analiz praktyk związanych z religią i praktyk dotyczących korzystania z programów opartych na technologii sztucznej inteligencji, można wykazać, że istnieje między nimi związek. Osoby niewykazujące aktywności religijnej przedstawiają większe zainteresowanie analizowaną technologią. Wśród osób praktykujących religię kilka razy w tygodniu 3,6% korzysta ze sztucznej inteligencji codziennie, 3,6% – kilka razy w tygodniu, 5,5% – raz w tygodniu, 9,1% – jeden, dwa razy w miesiącu, 34,5% – rzadziej niż raz w miesiącu, a 43,6% nie korzystało nigdy. U osób wierzących praktykujących raz w tygodniu odsetki korzystania ze sztucznej inteligencji wynoszą odpowiednio: 12,6%, 13,8%, 5,7%, 8,9%, 13,8%, 45,1%. W grupie praktykującej jeden, dwa razy w miesiącu odsetki korzystania ze sztucznej

inteligencji to: 11,4%, 25%, 9,1%, 10,2%, 21,6%, 22,7%. Osoby praktykujące jeden raz, kilka razy w roku korzystają ze sztucznej inteligencji w odsetkach kolejno: 12,1%, 13,7%, 4,8%, 12,9%, 12,1%, 44,4%. Praktykujący rzadziej niż raz w roku korzystają ze sztucznej inteligencji w odsetkach: 12,6%, 9,2%, 7,6%, 11,8%, 21,8%, 37%. Wśród respondentów, którzy nie uczestniczą w praktykach religijnych, odsetki korzystania ze sztucznej inteligencji wynoszą: 18,3%, 12%, 4,7%, 8%, 14,3%, 42,7%. Jednym z możliwych wyjaśnień jest to, że religia i technologie cyfrowe mogą być postrzegane jako systemy wartości konkurujące ze sobą. Religia często podkreśla znaczenie tradycyjnych wartości społecznych, takich jak rodzina, wspólnota, moralność, etyka. Nowe technologie mogą być postrzegane jako ich zagrożenie.

Analizując dane zawarte w tabeli 2, można zauważyć, że osoby praktykujące modlitwę kilka razy dziennie stanowią stosunkowo wysoki odsetek tych, które nigdy nie korzystały z technologii AI (40,7%). Na uwagę zasługuje fakt, iż bardzo wysoki odsetek osób nigdy niekorzystających z omawianych technologii znajduje się zarówno w każdej z kafeterii dotyczącej częstotliwości praktyk modlitwy codziennej, jak i wśród osób deklarujących brak praktyk modlitwy (44,8% badanych). Osoby praktykujące modlitwę raz dziennie charakteryzuje zróżnicowane podejście do sztucznej inteligencji. Najwięcej z nich (44,2%) zadeklarowało, że nie korzystało z tej technologii nigdy. W przypadku osób praktykujących modlitwę kilka razy w tygodniu odsetek niekorzystających ze sztucznej inteligencji wynosi 27,5% (podobny udział procentowy zaobserwowano wśród osób praktykujących modlitwę jeden, dwa razy w miesiącu), co jest wynikiem stosunkowo niskim w porównaniu do innych grup. Zaskakujące jest to, że zróżnicowanie danych jest zauważalne w grupie badanych, którzy korzystają z aplikacji bazujących na sztucznej inteligencji codziennie, i to zarówno wśród osób deklarujących codzienne praktyki modlitwy, jak i osób niepraktykujących w ogóle. Przykładowo różnica między osobami praktykującymi modlitwę raz dziennie a niepraktykującymi w ogóle wynosi 2 punkty procentowe, a w porównaniu z praktykującymi kilka razy dziennie – 3,5%. Podejmując próbę wyjaśnienia zaobserwowanej mało znaczącej różnicy, można założyć, że osoby praktykujące mogą wspierać się AI podczas modlitwy indywidualnej.

Celem uzupełnienia niniejszej części jest zaprezentowanie i omówienie danych statystycznych dotyczących praktyk modlitwy indywidualnej

(zmienna niezależna) i cech demograficznych (zmiennie zależne). Interpretacja ta stanowi wprowadzenie dotyczące religijności społeczeństwa polskiego, która sukcesywnie spada¹⁶. W tym wypadku skupiono się na praktykach modlitw indywidualnych, ponieważ w religijnych praktykach zauważalny jest znaczący spadek, zwłaszcza w młodej grupie Polaków (18–24 lata – 24,8% podkreśla, że są niewierzący¹⁷). Logiczne jest założenie, że wraz z odchodzeniem od praktyk religijnych, a także określeniem się jako niewierzący równoległe zanika praktyka modlitwy. Nasuwa się pytanie: Czy aby na pewno taki stan rzeczy istnieje? W społeczeństwie sieciowym zauważamy postępującą indywidualizację i anonimowość, której pęd nadaje cyfryzacja i wirtualność. I tu pod płaszczykiem anonimowości osoby mogą tworzyć trend indywidualnych (osobistych) praktyk modlitwy. „Poszerzenie tych możliwości wyboru działań – napisał Janusz Mariański – daje jednostce szansę wybierania spośród wielu opcji tych, które jej odpowiadają, a więc szansę określenia się w autonomii”¹⁸.

Tabela 3. Cechy religijne a praktyki modlitwy indywidualnej (dane w %)

		Praktyki modlitwy indywidualnej							Ogółem
		Kilka razy dziennie	Raz dziennie	Kilka razy w tygodniu	Raz w tygodniu	Jeden, dwa razy w miesiącu	Rzadziej niż jeden raz w miesiącu	Nie praktykuję modlitwy	
Płeć	Kobieta	11,1	16,6	10,7	6,0	10,6	10,4	34,6	100,0
	Mężczyzna	11,3	21,0	11,9	4,0	5,2	9,1	37,5	100,0
Wiek	15–24 lata	5,4	17,9	6,0	6,0	12,5	8,9	43,5	100,0
	25–34 lata	5,8	15,0	13,5	6,8	7,2	16,4	35,3	100,0
	35–44 lata	13,2	18,0	10,8	4,8	7,8	11,4	34,1	100,0
	45–54 lata	13,3	17,3	16,2	2,3	12,7	5,2	32,9	100,0
	55–65 lat	15,4	22,8	10,4	5,0	3,8	7,7	34,9	100,0

¹⁶ M. Grabowska, *Polski pejzaż religijny z daleka i z bliska*, w: *Sekularyzacja po polsku*, red. M. Grabowska, „Opinie i Diagnozy”, nr 50, CBOS, Warszawa 2022.

¹⁷ M. Grabowska, *Polski pejzaż religijny z dalekiego planu*, *Komunikat z badań*, CBOS, Warszawa 2022, s. 6.

¹⁸ J. Mariański, *Indywidualizacja religijna i moralna jako megatrend społeczno-kulturowy. Studium socjologiczne*, Wyd. Adam Marszałek, Toruń 2021, s. 71.

		Praktyki modlitwy indywidualnej							Ogółem
		Kilka razy dziennie	Raz dziennie	Kilka razy w tygodniu	Raz w tygodniu	Jeden, dwa razy w miesiącu	Rzadziej niż jeden raz w miesiącu	Nie praktykując modlitwy	
Wykształcenie	Podstawowe / gimnazjalne	22,2	14,8	7,4	3,7	7,4	14,8	29,6	100,0
	Zasadnicze	3,9	22,1	5,2	14,3	2,6	5,2	46,8	100,0
	Średnie	7,2	22,9	11,0	3,9	8,8	11,0	35,3	100,0
	Pomaturalne	4,4	12,6	12,6	6,7	11,1	9,6	43,0	100,0
	Wyższe licencjackie	16,8	15,9	7,1	4,4	11,5	8,8	35,4	100,0
Miejscowość zamieszkania	Wieś	13,2	22,2	14,3	5,3	5,8	8,5	30,7	100,0
	Małe miasto (do 20 tys. mieszkańców)	14,6	19,4	5,6	7,6	12,5	10,4	29,9	100,0
	Średnie miasto (od 20 do 99 tys. mieszkańców)	12,0	19,3	10,8	5,4	5,4	9,6	37,3	100,0
	Duże miasto (od 100 do 500 tys. mieszkańców)	8,5	17,9	12,4	4,4	9,2	9,8	37,9	100,0
	Wielkie miasto (pow. 500 tys. mieszkańców)	13,4	13,4	9,3	3,1	5,2	12,4	43,3	100,0

Źródło: opracowanie własne

Dane statystyczne zaprezentowane w tabeli 3 wskazują, iż modlitwy nie praktykuje większy odsetek mężczyzn w porównaniu do kobiet (37,5% mężczyzn do 34,6% kobiet). Kobiet praktykujących modlitwę raz dziennie jest natomiast więcej w porównaniu do mężczyzn (16,6% kobiet do 21,0% mężczyzn). Patrząc na dominującą częstotliwość modlitwy, najczęstszą wśród obu płci praktyką modlitewną jest modlitwa raz dziennie. Najwięcej mężczyzn (21,0%) oraz kobiet (16,6%) praktykuje modlitwę codziennie. Nieznaczne różnice są zauważalne, jeżeli chodzi o praktyki modlitwy codziennej rzadziej niż raz w tygodniu (10,4% kobiet i 9,1% mężczyzn). W obu grupach płciowych można zauważyć, że odsetek osób praktykujących modlitwę rośnie w miarę zwiększania się częstotliwości praktyk. Na przykład odsetek osób praktykujących modlitwę codziennie jest wyższy niż odsetek osób praktykujących modlitwę kilka razy w tygodniu czy kilka razy w miesiącu.

Należy podkreślić istotność statystyczną na poziomie 0,05 ($\chi^2 = 15,949$; $df = 6$; $p = 0,014$) między praktyką modlitwy indywidualnej a płcią.

Przedstawione dane, wyrażone procentowo, pokazują różnice między młodszymi a starszymi respondentami. W grupie wiekowej 15–24 lata prawie połowa (43,5%) deklaruje, że nie praktykuje modlitwy. Jednakże ci, którzy się modlą, robią to raczej sporadycznie – 17,9% raz dziennie i 5,4% kilka razy dziennie. Młodsze pokolenie jest mniej regularne w praktykowaniu modlitwy w porównaniu do starszych grup wiekowych. Podobny trend można zauważyć w grupie 25–34 lata, gdzie 35,3% respondentów nie praktykuje modlitwy, a jedynie 15% robi to codziennie. Taka tendencja może wskazywać na ewolucję nawyków duchowych w młodszych pokoleniach. Grupa wiekowa 35–44 lata prezentuje bardziej zróżnicowany obraz. Tutaj największa grupa, 34,1%, praktykuje modlitwę raz w tygodniu, co sugeruje pewną regularność. Jednakże 13,2% deklaruje modlitwę kilka razy dziennie, co oznacza większą różnorodność w nawykach duchowych w tej grupie wiekowej. W przypadku osób w wieku 45–54 lata 32,9% praktykuje modlitwę raz w miesiącu. Jednak znaczna liczba respondentów, 17,3%, modli się codziennie, co stanowi wyraźne odróżnienie od młodszych grup wiekowych. Najstarsza grupa wiekowa, 55–65 lat, prezentuje najwyższą skłonność do regularnej modlitwy. 34,9% respondentów modli się raz w tygodniu, a 22,8% codziennie. To sugeruje, że z wiekiem wzrasta tendencja do regularnego praktykowania modlitwy indywidualnej. Analizy danych wskazują na zmienność praktyk modlitewnych w różnych grupach wiekowych. Młodsze pokolenia wydają się mniej zaangażowane w regularne modlitwy niż starsze, ale istnieją także zauważalne zróżnicowania w nawykach duchowych w obrębie poszczególnych grup wiekowych.

Badania dotyczące praktyk modlitwy indywidualnej wśród osób o różnym poziomie wykształcenia wykazują, iż 29,6% respondentów z wykształceniem podstawowym lub gimnazjalnym nie praktykuje modlitwy, co stanowi największy odsetek w tej kategorii. Jednak aż 22,2% osób z tego samego poziomu edukacji praktykuje modlitwę kilka razy dziennie, co może świadczyć o zróżnicowaniu praktyk religijnych w tej grupie.

W przypadku osób posiadających wykształcenie zasadnicze prawie połowa z nich, 46,8%, nie praktykuje modlitwy, co jest najwyższym odsetkiem w porównaniu do innych poziomów edukacji. Z tej grupy 22,1% modli się natomiast raz dziennie. Zauważa się, że wraz ze wzrostem poziomu wykształcenia maleje

odsetek osób niepraktykujących modlitwy, wzrasta jednocześnie liczba tych, którzy modlą się codziennie. Osoby ze średnim wykształceniem wykazują bardziej zrównoważone podejście do praktyk modlitewnych, gdzie żadna z kategorii nie dominuje znacząco, jednak 35,3% z nich nie praktykuje modlitwy, co jest odsetkiem niższym niż w przypadku grupy z wykształceniem zasadniczym. Ciekawe jest również spojrzenie na osoby z wykształceniem wyższym licencjackim. W tej grupie 35,4% respondentów nie praktykuje modlitwy, a największy odsetek, 16,8%, modli się kilka razy dziennie. Uzyskane dane sugerują, że istnieje związek między poziomem wykształcenia a praktykami modlitewnymi. Niższy poziom edukacji często koreluje z mniejszą skłonnością do praktyk religijnych, zwłaszcza jeśli chodzi o modlitwy regularne. Wraz ze wzrostem poziomu wykształcenia rośnie natomiast zróżnicowanie w praktykach modlitewnych, obejmujące osoby zarówno praktykujące regularnie, jak i te, które nie praktykują wcale.

Najwyższy odsetek osób niepraktykujących modlitwy występuje wśród mieszkańców dużych miast powyżej 500 tys. mieszkańców, gdzie aż 43,3% respondentów nie praktykuje modlitwy. Jednocześnie mieszkańcy tych obszarów wykazują mniejszą skłonność do codziennych praktyk modlitewnych w porównaniu do innych grup. Na drugim końcu spektrum znajdują się mieszkańcy wsi, którzy charakteryzują się największym odsetkiem osób deklarujących praktyki modlitwy kilka razy dziennie i raz dziennie (łącznie 34%). Jednak i tutaj odsetek osób niepraktykujących modlitwy jest znaczący i wynosi 30,7%. Małe miasta (do 20 tys. mieszkańców) wydają się charakteryzować umiarkowanym poziomem praktyk modlitewnych, gdzie 29,9% nie praktykuje modlitwy, a 19,4% modli się codziennie. Mieszkańcy średnich miast (20–99 tys. mieszkańców) wykazują wyższy odsetek praktyk modlitewnych, zwłaszcza kilka razy w tygodniu (10,8%), co sugeruje większe zaangażowanie w te praktyki w porównaniu do mieszkańców dużych miast.

Po wnikliwym przeanalizowaniu materiału empirycznego rodzi się konkluzja, że młodsze pokolenia są mniej religijne ze względu na zmianę wartości i priorytetów. Młodzi ludzie są bardziej otwarci na świat i bardziej skłonni do poszukiwania własnej drogi duchowej. Mogą oni również kwestionować tradycyjne wartości i autorytety, w tym religię. Niższy poziom wykształcenia może być również związany z mniejszą skłonnością do praktyk religijnych. Osoby z niższym wykształceniem są mniej podatne na wpływ religii, zwłaszcza w okresie dorastania i wczesnej dorosłości. Mogą one również mieć mniej czasu i energii

na praktyki religijne, ponieważ są bardziej skupione na rozwoju zawodowym i rodzinnym. Na praktyki religijne może wpływać również zamieszkanie w dużym mieście. Duże miasta często są pod względem religijnym bardziej zróżnicowane, co może utrudniać praktykowanie jednej religii. Ponadto mieszkańcy takich miast są często bardziej zajęci i mają mniej czasu na praktyki religijne.

Definicje przedmiotowe religii

Definicje przedmiotowe ujmują religię ze względu na obiekt czci religijnej i wiary. Tu należy podkreślić, iż pojęcie religii rozpatruje się w znaczeniu przedmiotowym lub podmiotowym. Stanisław Kuczkowski trafnie wyjaśnił istotę podziału, gdzie przez religię rozumie się zbiór prawd, nakazów i zakazów regulujących stosunek między człowiekiem i Bogiem. „Pojęcie religijność [...] określa podmiotowe, subiektywne, indywidualne ustosunkowanie się osoby do systemu prawd, przy zaangażowaniu całej osobowości”¹⁹. W celu ustalenia przedmiotu/przedmiotów kultu we współczesnej kulturze algorytmicznej przedstawiono respondentom listę potencjalnych przedmiotów kultu z prośbą o wskazanie tych, o których myślą najczęściej.

Tabela 4. Przedmioty kultu we współczesnej kulturze algorytmicznej

Przedmioty kultu	Liczba wskazanych odpowiedzi (N)	Procent wskazanych odpowiedzi (%)
Bóg	521	70,4
Istota Wyższa	169	22,9
Siły Wszechświata/Kosmosu	184	24,9
Dobra materialne	113	15,3
Rzeczy obdarzone mocą (np. tarot, horoskopy, wahadélka itp.)	60	8,1
Nowe technologie (np. sztuczna inteligencja)	46	6,2
Nic	41	5,5
Ogółem	1134	153,4

Uwaga: Dane nie sumują się do 100%, ponieważ respondent mógł wybrać więcej niż jedną odpowiedź.

Źródło: opracowanie własne

¹⁹ S. Kuczkowski, *Psychologia religii*, Wydawnictwo WAM, Kraków 1998, s. 22.

We współczesnej kulturze algorytmicznej najczęściej wskazywanym przez badanych Polaków czynnikiem modelującym przedmiot kultu jest Bóg (521 odpowiedzi; 70,4%). Na przeciwnym biegunie badani usytuowali nowe technologie, m.in. sztuczną inteligencję (6,2% odpowiedzi). Nie jest zaskoczeniem, że odpowiedzi o największej liczbie wskazań dotyczą sił metafizycznych oscylujących wokół materii boskiej. Wskaźnik ten obrazuje, że mimo Castellsowskiego społeczeństwa sieciowego²⁰ istnieje w polskiej społeczności tradycyjny obraz wyobrażeń na temat sił nadprzyrodzonych.

W dalszej części analiz przeprowadzono studium porównawcze danych statystycznych opartych na częstotliwości praktyk z wykorzystaniem programów bazujących na technologii sztucznej inteligencji oraz praktyki modlitwy indywidualnej (tabela 5). Z danych wynika, iż modlitwę indywidualną codziennie praktykuje 29,7% respondentów, podczas gdy z programów AI korzysta jedynie 13,5% z nich. Minimalna przewaga modlitwy pojawia się również w praktykach kilka razy w tygodniu (11,3% dla modlitwy i 13,1% dla SI).

Jeśli chodzi o częstotliwość praktykowania raz w tygodniu, 5,7% respondentów korzysta z narzędzi opartych na AI, podczas gdy 5,0% decyduje się na modlitwę. Praktykę korzystania z AI jeden, dwa razy w miesiącu podejmuje 10,1% respondentów, a 8,2% podejmuje praktykę modlitwy. Rzadkie korzystanie z praktyk (mniej niż raz w miesiącu) jest natomiast bardziej powszechne w przypadku AI (16,2%) niż modlitwy (9,8%). Należy zauważyć, że istnieje znaczna grupa respondentów (41,4%), która nie korzystała z programów AI nigdy, podczas gdy 36,0% nigdy nie praktykuje modlitwy. Sugeruje to, że choć modlitwa indywidualna jest bardziej powszechna, to istnieje też duży odsetek osób, które nie mają doświadczenia z technologią opartą na AI.

²⁰ M. Castells, *Spoleczeństwo sieci*, PWN, Warszawa 2007.

Tabela 5. Praktyki z programami bazującymi na technologii sztucznej inteligencji i praktyki z modlitwą indywidualną

Częstotliwość praktyk	Praktyki z programami bazującymi na sztucznej inteligencji <i>N</i> = 1052		Praktyki z modlitwą indywidualną <i>N</i> = 1052		
	Częstość	Procent	Częstość	Procent	
Codziennie	142	13,5	26,6	314	29,7
Kilka razy w tygodniu	137	13,1		119	11,3
Raz w tygodniu	60	5,7	53	5,0	
Jeden, dwa razy w miesiącu	106	10,1	84	8,2	
Rzadziej niż jeden raz w miesiącu	171	16,2	103	9,8	
Nigdy nie korzystałem z SI/ Nie praktykuję modlitwy	436	41,4	379	36,0	
Ogółem	1052	100,0	1052	100,00	

Źródło: opracowanie własne

Różnice w częstotliwości praktyk odzwierciedlają złożoność podejścia do tych dwóch obszarów życia. Częstotliwość angażowania się w te praktyki zależy od wielu czynników, w tym osobistych nawyków, przekonań religijnych i wygody korzystania z technologii. Studium to rzuca światło na zróżnicowane podejście do praktyk opartych na technologii w porównaniu z praktykami religijnymi, a jednocześnie podkreśla, że obie te dziedziny mają swoje miejsce w życiu społecznym. Badania pokazują, że ludzie, którzy mają regularne nawyki związane z technologią, są bardziej skłonni do częstego korzystania z technologii w celach religijnych. Dzieje się tak dlatego, że są już przyzwyczajeni do korzystania z technologii i czują się w niej komfortowo. Podobnie ludzie o silnych przekonaniach religijnych są bardziej skłonni do angażowania się w praktyki religijne. Poświęcają temu czas, ponieważ religia jest dla nich ważną częścią życia. Oczywiście nie wszyscy ludzie podchodzą do praktyk opartych na technologii i praktyk religijnych w taki sam sposób. Istnieją pewne czynniki środowiskowe, które również mogą wpływać na te wybory. Na przykład ludzie żyjący w społecznościach o silnych tradycjach religijnych są bardziej skłonni do regularnego uczestnictwa w praktykach religijnych.

Definicje funkcjonalne

W definicjach funkcjonalnych religii przedmiot wiary oraz czci jest bez znaczenia. Funkcjoniści, ujmując religię jako pozytywną siłę w życiu społeczeństwa, definiują ją ze względu na pełnione przez nią funkcje²¹. Émile Durkheim wskazuje na integracyjne, psychologiczne oraz egzystencjalne funkcje religii²². Według Normana Goodmana religia nadaje sens życiu i wsparcie psychiczne, odpowiada za spójność społeczeństwa i kontrolę społeczną²³. Religia, spełniając ważne funkcje społeczne, istnieje w jakiejś formie we wszystkich społeczeństwach, także we współczesnym społeczeństwie algorytmicznym.

Thomas Luckmann jest autorem typowo funkcjonalnej definicji religii. W jego rozważaniach można odnaleźć tożsamość z tym, co głęboko ludzkie. „Funkcjonalna definicja (...) stawia sobie za cel ogarnięcie możliwie szerokiego spektrum treści wierzeń i społecznych form, które wypełniają funkcje «religijne», nie chcąc zawęzić pola widzenia do tego, co tradycyjne instytucje religijne określają jako religię”²⁴. Rodzą się więc kolejne pytania: Które z funkcji religii są realizowane przez technologie oparte na sztucznej inteligencji? Analizie porównawczej zostaną poddane wybrane funkcje.

Funkcja kontroli społecznej: podejmowanie decyzji życiowych

Każdy system religijny posiada własne dogmaty, czyli podstawowe tezy podane do wierzenia jako nienaruszalna prawda, pewnik, przekonania²⁵. Tezy te powinny być postrzegane także przez wyznawców danej religii jako

²¹ Zwolennicy teorii konfliktu nie podzielają pozytywnego zdania funkcjonalistów o religii jako instytucji społecznej. Wskazują na kilka negatywnych cech społeczeństwa, które wynikają bezpośrednio z religii albo są przez nie utrwalane. Zob. N. Goodman, *Wstęp do socjologii*, dz. cyt., s. 217–219.

²² É. Durkheim, *Elementarne formy życia religijnego*, tłum. A. Zadrożyńska, PWN, Warszawa 2010, s. 20.

²³ N. Goodman, *Wstęp do socjologii*, dz. cyt., s. 217–219.

²⁴ H. Knoblauch, „Niewidzialna religia” Thomasa Luckmanna, czyli o przemianie religii w religijność, w: Th. Luckmann, *Niewidzialna religia. Problem religii we współczesnym społeczeństwie*, tłum. L. Bluszcz, Zakład Wydawniczy „Nomos”, Kraków 1996, s. 13; zob. N. Luhmann, *Funkcja religii*, tłum. D. Motak, Zakład Wydawniczy „Nomos”, Kraków 1998, s. 11–72.

²⁵ *Dogmat*, w: *Słownik języka polskiego*, PWN, <http://sjp.pl/dogmat> (3.01.2024).

wyznaczniki/punkty odniesienia przy podejmowaniu decyzji życiowych, a także budowaniu poczucia własnej skuteczności bazującego na tych decyzjach. Mariola Chomczyńska-Rubacha i Krzysztof Rubacha przyjmują, „że poczucie własnej skuteczności powstaje na bazie osobistych doświadczeń jednostki w zakresie dysponowania zasobami motywacyjnymi i poznawczo-działaniowymi. Na te pierwsze składają się: umiejętność odrzucania gratyfikacji, wiara we własne siły, motywacja rozwojowa i wytrwałość w działaniu. Zasoby poznawczo-działaniowe budują natomiast takie zmienne jak: umiejętność przekładania celów na program działania, odporność na frustrację i stres, poczucie sprawstwa oraz wewnątrzsterowność”²⁶. Stąd kolejne pytanie: Czy modlitwa i sztuczna inteligencja wpływają na decyzje życiowe badanych osób?

Tabela 6. Podejmowanie decyzji życiowych (dane w %)

	Czy modlitwa wpływa na Pani/Pana podejmowanie decyzji życiowych? N = 1052	Czy AI (np. chatboty, personalizacja treści) wpływa na Pani/Pana podejmowanie decyzji życiowych? N = 608
Tak, często opieram na niej swoje decyzje.	17,9	4,1
Tak, ale czasem także konsultuję się z innymi źródłami wsparcia.	19,7	28,6
Nie, modlitwa/AI nie wpływa na moje decyzje.	62,4	67,3
Ogółem	100,0	100,0

Źródło: opracowanie własne

Z uzyskanych danych wynika, że dla większości badanych Polaków ani modlitwa (62,4%), ani sztuczna inteligencja (67,3%; treści wygenerowane przez AI, personalizacja treści) nie wpływają na podejmowane decyzje życiowe. Wynik jest porównywalny; różnica oscyluje wokół 5 punktów procentowych. Wynik ten implikuje pytanie o brak spójności między nauczaniem Kościoła i życiem codziennym. Jednocześnie wskazuje na większe zaufanie

²⁶ M. Chomczyńska-Rubacha, K. Rubacha, *Test poczucia skuteczności. Podręcznik Pracowni Narzędzi Badawczych Komitetu Nauk Pedagogicznych PAN*, Wydawnictwo Naukowe UMK, Warszawa–Toruń 2013, s. 7–8.

respondentów do wyników procesu algorytmicznego niż do wiedzy i intuicji człowieka i religii. Uzyskane dane korespondują z opisanym w literaturze przedmiotu pojęciem autorytetu algorytmicznego²⁷.

Warto się zastanowić nad przyczynami owego braku spójności między nauczaniem Kościoła a codziennym życiem. Może to wynikać z ewolucji społeczeństwa, zmiany wartości czy indywidualnych doświadczeń. Szczególnie istotne jest, że większość respondentów wykazuje większe zaufanie do wyników algorytmów niż do ludzkiej wiedzy i intuicji. Może to świadczyć o postrzeganiu algorytmów jako obiektywnych i pozbawionych emocji w porównaniu do ludzi.

Zjawisko to koresponduje z koncepcją **autorytetu algorytmicznego**, co oznacza przekonanie o rzetelności decyzji podejmowanych przez algorytmy na podstawie danych, bez wpływu emocji czy uprzedzeń. Implikacje tego wyniku dla społeczeństwa i instytucji religijnych są istotne. Współczesny człowiek może bardziej ufać technologii, poszukując obiektywności i przewidywalności w podejmowaniu decyzji życiowych. Kościół może potrzebować dostosować swoje nauczanie do zmieniających się wartości i oczekiwań społeczeństwa, aby utrzymać spójność z życiem codziennym wiernych.

Funkcje psychologiczne: wpływ na samopoczucie

Émile Durkheim²⁸, Norman Goodman²⁹ i inni wskazują na psychologiczne funkcje religii. Prawidłowo przeżywana religijność jest źródłem pokoju wewnętrznego. Stąd kolejne pytanie: W jakim stopniu modlitwa i korzystanie z technologii opartych na sztucznej inteligencji wpływają na samopoczucie respondentów?

²⁷ J. Kreft, *Władza platform. Za fasadą Google, Facebooka i Spotify*, Universitas, Kraków 2021, s. 29.

²⁸ É. Durkheim, *Elementarne formy życia...*, dz. cyt., s. 20.

²⁹ N. Goodman, *Wstęp do socjologii*, dz. cyt., s. 217–219.

Tabela 7. Wpływ na samopoczucie

Skala	W jakim stopniu modlitwa wpływa na Pani/Pana samopoczucie? N = 1052 (dane %)		W jakim stopniu korzystanie z technologii opartych na sztucznej inteligencji wpływa na Pani/Pana samopoczucie? N = 608 (dane %)	
Zdecydowanie pozytywnie	24,0	53,8	9,0	41,0
Raczej pozytywnie	29,8		32,0	
Raczej negatywnie	7,0	10,7	4,8	7,8
Zdecydowanie negatywnie	3,7		3,0	
Trudno powiedzieć	41,8		51,2	
Ogółem	100,0		100,0	

Źródło: opracowanie własne

Analiza tabeli nasuwa następujące wnioski: istnieje wysoki wskaźnik procentowy respondentów udzielających odpowiedzi, że zarówno modlitwa (53,8%), jak i korzystanie z technologii opartych na AI (41%) wpływa na ich samopoczucie pozytywnie, oraz wysoki odsetek badanych mających trudności z jednoznaczną odpowiedzią na zadane pytanie. Jednocześnie 75,8% respondentów korzysta ze sztucznej inteligencji w celu rozwiązania codziennych problemów; 28% jest przekonanych o nieograniczonych możliwościach AI; 9,4% korzysta z AI w celu budowania/podtrzymywania relacji z wirtualnym, niewidzialnym, niedoświadczalnym światem. Zastanawia ponad 10% wynik negatywnego wpływu modlitwy na samopoczucie Polaków, przy porównywalnym wyniku (8%) dla poszukiwania pocieszenia i wsparcia w technologii AI.

Funkcje egzystencjalne

Na wstępie należy się pochylić nad samym egzystencjalizmem, w przypadku którego można mówić o filozofii podmiotu mającej „charakter antynaturalistyczny, antypozytywistyczny i antyracjonalistyczny, która koncentruje się na konkretnym ludzkim bycie zawsze doświadczanym przez jednostkę subiektywnie, wewnątrznie, na zagadnieniach związanych z wolnością, samotnością i skończonością człowieka w świecie, problemach ujmowanych w kategoriach dramatyizmu ludzkiego istnienia”³⁰.

³⁰ D. Bajer, *Egzystencjalizm jako filozofia podmiotu*, „IDEA – Studia nad strukturą i rozwojem pojęć filozoficznych” 2011, t. 23, s. 116.

Niniejszy rozdział, obok aspektów religijnych, podejmuje również aspekt technologiczny. Analizując literaturę przedmiotu, zauważyła się, że nie ma zbyt wielu badań empirycznych i analiz teoretycznych dotyczących samej egzystencjalności człowieka i sztucznej inteligencji czy też humanoidalnych wytworów ludzkiego umysłu i „ręki”. W celu określenia opinii związanej z egzystencjalną funkcją AI zastosowano serię pytań, w stosunku do których badani, na 5-stopniowej skali, wyrażali własne zdanie na temat sztucznej inteligencji.

Z analiz wynika, iż w odpowiedziach przeważały punkty 3 i 4 (*ani się zgadzam, ani się nie zgadzam i raczej się zgadzam*). 30,6% badanych uznaje, że zazwyczaj nie wie, kiedy korzysta z technologii sztucznej inteligencji, a 30,1% nie ma zdania w tej kwestii, wskazując na skali odpowiedź nr 3. Z danych wynika, iż 26,4% postrzega sztuczną inteligencję jako zagrożenie, a 31% uznaje, że jest ona zdecydowanie bardziej niebezpieczna niż inne wytwory techniczne. 31% respondentów uznaje, że AI to narzędzie do sprawowania kontroli nad ludźmi, 28,6% ani się nie zgadza, ani się z tym zgadza. 28% uczestników badania ma poczucie, że sztucznej inteligencji nikt nie kontroluje. Poczucie, że sztuczna inteligencja za bardzo ingeruje w życie ludzi, zadeklarowało 20,6% respondentów (raczej się z tym zgadza), a 19,9% raczej się nie zgadza (36% ani się zgadza, ani się nie zgadza). Jednak można się doszukiwać nie tylko negatywnych elementów egzystencjalnych związanych z AI, przyczyniających się do budowy teorii chaosu i/lub konfliktu. 49,7% badanych uznaje, że sztuczna inteligencja usprawni wiele aspektów życia, a 15,4% raczej się z tym stwierdzeniem zgadza. Zdaniem 65,9% uczestników badania (raczej się zgadza i zdecydowanie się zgadza) AI stanowi dla ludzi odkrycie przełomowe.

Tabela 8. Opinia na temat AI w społeczeństwie (skala 1–5); N = 1052 (dane w %)

1 – Zdecydowanie się nie zgadzam; 2 – Raczej się nie zgadzam; 3 – Ani się zgadzam, ani się nie zgadzam; 4 – Raczej się zgadzam; 5 – Zdecydowanie się zgadzam; Tp. – Trudno powiedzieć.

	1	2	3	4	5	Tp.
Zazwyczaj nie wiem, kiedy korzystam z technologii sztucznej inteligencji.	7,1	21,7	30,1	30,6	7,1	3,3
Generalnie postrzegam sztuczną inteligencję jako zagrożenie.	8,3	19,9	32,5	26,4	9,7	3,2
Odczuwam niepokój związany z rozwojem sztucznej inteligencji.	9,3	18,0	27,5	30,1	12,2	3,0
Sztuczna inteligencja jest zdecydowanie bardziej niebezpieczna niż inne wytwory techniczne.	6,7	16,7	28,6	31,0	13,0	4,0

	1	2	3	4	5	Tp.
Sztuczna inteligencja to narzędzie do sprawowania kontroli nad ludźmi.	9,1	13,3	28,6	30,1	15,1	3,8
Mam poczucie, że nikt nie kontroluje sztucznej inteligencji.	6,9	14,8	33,0	28,0	12,6	4,5
Mam poczucie, że sztuczna inteligencja za bardzo ingeruje w moje życie.	11,4	19,9	36,0	20,6	9,1	3,1
Sztuczna inteligencja usprawni wiele aspektów życia.	2,9	5,6	24,5	49,7	15,4	1,8
Sztuczna inteligencja to przełomowe odkrycie dla ludzi.	2,6	5,2	24,3	45,0	20,9	2,0
Uważam, że sztuczna inteligencja może dawać wartościowe odpowiedzi na pytania religijno-egzystencjalne.	16,1	19,3	33,4	17,4	6,0	7,8
Sztuczna inteligencja może stać się obiektem kultu podobnego do kultu religijnego.	15,2	23,6	27,7	19,4	7,0	7,0
Używanie sztucznej inteligencji może zastąpić praktyki religijne.	23,3	29,2	29,5	8,0	3,1	6,8
Tworzenie tekstu modlitwy za pomocą sztucznej inteligencji uważam za profanację.	11,0	13,7	29,3	19,9	19,7	6,5
Tworzenie grafik/obrazów/zdjęć religijnych za pomocą sztucznej inteligencji uważam za profanację.	12,5	18,1	34,6	16,1	13,2	5,6
Traktuję sztuczna inteligencję jako coś bardzo ważnego w moim życiu.	27,5	28,3	30,7	8,2	2,4	2,9
Bardziej wierzę w możliwości sztucznej inteligencji niż możliwości Boga.	39,3	15,2	22,6	8,6	7,6	6,8

Źródło: opracowanie własne

Na skali zauważa się zwiększenie udziału procentowego odpowiedzi przeczących (*raczej się nie zgadzam* i *zdecydowanie się nie zgadzam*) tam, gdzie padają pytania związane z religią, a zwłaszcza gdzie zmierzają w kierunku określenia AI jako *sacrum*, które mogłoby zastąpić w społeczeństwie Boga. Postawy obojętności (*ani się zgadzam, ani się nie zgadzam*) oscylują w granicach rzędu 30%. Łącznie 35,4% badanych nie zgadza się (*zdecydowanie się nie zgadza* i *raczej się nie zgadza*), że sztuczna inteligencja może dawać wartościowe odpowiedzi na pytania religijno-egzystencjalne; 38% nie zgadza się, że używanie sztucznej inteligencji może zastąpić praktyki religijne; 55,8% nie traktuje AI jako czegoś bardzo ważnego w swoim życiu, a 54,5% nie zgadza się z twierdzeniem, iż bardziej wierzy w możliwości sztucznej inteligencji niż możliwości Boga. Zaskakujące wyniki dotyczą tworzenia tekstu modlitwy za pomocą sztucznej inteligencji, ponieważ 39,6% (*raczej się zgadza/zdecydowanie się zgadza*) uważa to za profanację, przy głosach przeciwnych 24,7% badanych. Podobnie jest z wynikami, jeżeli chodzi o tworzenie za pomocą sztucznej inteligencji grafik/obrazów/zdjęć religijnych, które uznaje się za profanację (29,3% zgadza się, 29,0% przyjmuje postawy neutralne, 30,6% nie zgadza się).

Powyższe wyniki badań empirycznych ukazują zróżnicowane podejścia i opinie społeczeństwa w temacie sztucznej inteligencji oraz jej potencjalnych implikacji egzystencjalnych. Zauważa się, że większość respondentów utrzymuje neutralne stanowisko (punkty 3) w kwestii korzystania z technologii sztucznej inteligencji, co może sugerować brak jednoznacznej oceny wpływu AI na ich życie. Warto podkreślić, że istnieje znaczna niepewność co do świadomości użytkowników w kontekście korzystania z technologii sztucznej inteligencji, ponad 30% respondentów deklaruje bowiem, że zazwyczaj nie zdaje sobie sprawy, kiedy używa AI. Jednocześnie istnieje obawa związana z postrzeganiem sztucznej inteligencji jako potencjalnego zagrożenia; ponad jedna czwarta grupy badanej obawia się negatywnych skutków związanych z jej rozwinięciem. Niezwykle istotne są również wyniki dotyczące aspektów kontrolowania ludzi przez sztuczną inteligencję, pokazujące, że ponad 30% respondentów uznaje AI za narzędzie do sprawowania kontroli nad ludźmi. To stanowisko może odzwierciedlać obawy związane z potencjalnym nadużyciem AI w zakresie prywatności i autonomii jednostki. Wyniki dotyczące religii w społeczeństwie polskim nie stanowią zaskoczenia. Większość respondentów nie zgadza się, że AI może dawać wartościowe odpowiedzi na pytania religijno-egzystencjalne, a także nie akceptuje możliwości zastąpienia praktyk religijnych przez korzystanie z tej technologii. Badania sugerują, że społeczeństwo ma zróżnicowane podejście do sztucznej inteligencji, obejmujące zarówno entuzjazm w odniesieniu do jej potencjalnych korzyści, jak i obawy związane z etycznymi i egzystencjalnymi aspektami jej rozwoju.

Podsumowanie i wnioski

Na wstępie procesu badawczego przyjęto hipotezę, że sztuczna inteligencja może się stać substytutem religii współczesnego człowieka. Stąd głównym celem artykułu było jej zweryfikowanie przez analizę wyników badań empirycznych przeprowadzonych wśród 1052 Polek i Polaków.

Za potwierdzeniem hipotezy przemawiają dane wpisujące się w ujęcie przedmiotowe definicji religii. Pomimo wzrastającej sekularyzacji społeczeństwa polskiego ponad 70% badanych wskazało na Boga (521 odpowiedzi). Nowe technologie (w tym sztuczna inteligencja) uzyskały wynik oscylujący

wokół 6% (46 odpowiedzi). Jednocześnie respondenci bardziej wierzą w możliwości AI niż możliwości Boga (2,30 w skali 1–5). Według badanych sztuczna inteligencja może się stać obiektem kultu podobnego do kultu religijnego: 2,74; może także zastąpić praktyki religijne: 2,39.

Definicje funkcjonalne ujmują religię w perspektywie pełnionych przez nią funkcji. Sztuczna inteligencja zdaniem wielu badanych nadaje sens życiu, traktują ją jako jego ważny aspekt (2,23 w skali 1–5); religia udziela wsparcia psychicznego (5,7%); pozytywnie wpływa na ich samopoczucie (41%). W opinii badanych może też dawać wartościowe odpowiedzi na pytania religijno-egzystencjalne (2,74). Sprawuje ponadto kontrolę społeczną nad ludźmi (3,50) oraz pełni funkcję integracyjną: budowania i/lub podtrzymywania relacji z wirtualnym, niewidzialnym, niedoświadczalnym światem (9,4%).

Należy podkreślić, że chociaż technologia sztucznej inteligencji staje się substytutem przedmiotu kultu, czci i wyznania i spełnia niektóre funkcje religijne, to nigdy nie stanie się religią. Nie uwzględnia ona całościowej koncepcji człowieka. Odchodzi zarówno od chrześcijańskiej wizji człowieka, jak i filozofii klasycznej, akcentującej harmonię ducha i ciała. W problematyce sztucznej inteligencji pomijany jest problem duszy, duchowości i sfery religijnej ucieleśnionej AI, m.in. w formie robotów humanoidalnych. Powód jest prozaiczny, a wyjaśnienie oparte na aspektach technologiczno-programistycznych. Sztuczna inteligencja jest wytworem ludzkim, opracowanym przez programistów, działającym na zasadach algorytmicznych (schematyczność, powtarzalność). Wytwór informatyczny może człowiekowi jedynie pomóc w wielu aspektach życia, tak codziennego (*profanum*), jak i duchowego (*sacrum*)³¹. Należy podkreślić, że nie jest on w stanie człowieka zastąpić, ponieważ nie ma świadomości. Działania wykonywane przez AI mogą sprawiać tylko wrażenie świadomych. Podobnie nie musi być świadome użytkowanie sztucznej inteligencji. W religii chrześcijańskiej, zakładającej antropocentryzm, podmiot społeczny ma naturę dialogiczną, tworzy przestrzeń relacji komunikacyjnych, włącznie z możliwością doświadczenia drugiego człowieka. Religia zakłada relacyjność, budowanie wspólnoty. Komunikacyjność sztucznej inteligencji ogranicza się natomiast tylko do interaktywności

³¹ Zob. M. Gruchoła, M. Sławek-Czochra, R. Zieliński, *Artificial Intelligence as a Tool Supporting Prayer Practices*, „Religions” 2024, vol. 15, no. 3, s. 271, <https://doi.org/10.3390/rel15030271>.

i sprawczości, czyli możliwości wzajemnego oddziaływania na siebie różnych systemów oraz układów urządzeń, które mogą wymieniać się informacjami i modyfikować wzajemnie swoje działania. Brakuje dialogu, wzajemnej komunikacji, kształtowania i rozumienia znaczeń. Hipoteza została zweryfikowana pozytywnie. Sztuczna inteligencja może stać się substytutem religii współczesnego człowieka – jedynie substytutem.

Bibliografia

- Bajer D., *Egzystencjalizm jako filozofia podmiotu*, „IDEA – Studia nad strukturą i rozwojem pojęć filozoficznych” 2011, t. 23, s. 115–134.
- Castells M., *Spoleczeństwo sieci*, PWN, Warszawa 2007.
- Chojnowski M., *Etyka sztucznej inteligencji*, Centrum Etyki Technologii Instytutu Humanities, Warszawa 2022.
- Chomczyńska-Rubacha M., Rubacha K., *Test poczucia skuteczności. Podręcznik Pracowni Narzędzi Badawczych Komitetu Nauk Pedagogicznych PAN*, Wydawnictwo Naukowe UMK, Warszawa–Toruń 2013.
- Dogmat*, w: *Słownik języka polskiego*, <http://sjp.pl/dogmat> (3.01.2024).
- Durkheim É., *Elementarne formy życia religijnego*, tłum. A. Zadrozynańska, PWN, Warszawa 2010.
- Główny Urząd Statystyczny, *Wyznania religijne w Polsce w latach 2019–2021. Analizy statystyczne*, GUS, Warszawa 2022.
- Goodman N., *Wstęp do socjologii*, tłum. J. Polak, J. Raszkowski, U. Zielińska, Zysk i S-ka, Poznań 2009.
- Grabowska M., *Polski pejzaż religijny z daleka i z bliska*, w: *Sekularyzacja po polsku*, red. M. Grabowska, „Opinie i Diagnozy”, nr 50, CBOS, Warszawa 2022.
- Grabowska M., *Polski pejzaż religijny z dalekiego planu. Komunikat z badań*, CBOS, Warszawa 2022.
- Gruchoła M., *Konsumpcjonizm – nową religią współczesnego człowieka?*, „Roczniki Kulturoznawcze” 2017, t. 8, nr 3, s. 39–62.
- Gruchoła M., Sławek-Czochra M., Zieliński R., *Artificial Intelligence as a Tool Supporting Prayer Practices*, „Religions” 2024, vol. 15, no. 3, s. 271, <https://doi.org/10.3390/rel15030271>.
- Gruchoła M., Zieliński R., Zajic M., Koper J., Popielewicz-Durakiewicz M., *Badanie postaw Polek i Polaków wobec AI*, Zakład Edukacji Medialnej i Sztucznej Inteligencji IBE, Warszawa 2023.
- Knoblauch H., *„Niewidzialna religia” Thomasa Luckmanna, czyli o przemianie religii w religijność*, tłum. D. Motak, w: Th. Luckmann, *Niewidzialna religia. Problem religii we współczesnym społeczeństwie*, tłum. L. Bluszcz, Zakład Wydawniczy „Nomos”, Kraków 1996, s. 7–41.

- Kreft J., *Władza platform. Za fasadą Google, Facebooka i Spotify*, Universitas, Kraków 2021.
- Kuczkowski S., *Psychologia religii*, Wydawnictwo WAM, Kraków 1998.
- Kurp F., *Sztuczna inteligencja od podstaw*, Helion, Gliwice 2023.
- Luhmann N., *Funkcja religii*, tłum. D. Motak, Zakład Wydawniczy „Nomos”, Kraków 1998.
- Mariański J., *Indywidualizacja religijna i moralna jako megatrend społeczno-kulturowy. Studium socjologiczne*, Wyd. Adam Marszałek, Toruń 2021.
- Nowakowski M., *O moralnej odpowiedzialności HAL-a 90001, czyli etyka sztucznej inteligencji w praktyce. Czy potrzebujemy definicji sztucznej inteligencji?*, „Prawo Mediów Elektronicznych” 2022, nr 1, s. 4–10.
- Pawluczuk W., *Religia*, w: *Encyklopedia socjologii*, red. Z. Boksański, t. 3, Oficyna Naukowa, Warszawa 2000, s. 288–289.
- Przybysz M., *Nie znam człowieka nawróconego dzięki sztucznej inteligencji. Ale może to kwestia czasu?*, <https://pl.aletia.org/2023/01/21/sztuczna-inteligencja-w-kosciele/> (2.01.2024).
- Religia*, w: *Encyklopedia PWN*, <http://encyklopedia.pwn.pl/haslo/religia;3966983.html> (28.12.2023).
- Różanowski K., *Sztuczna inteligencja: rozwój, szanse i zagrożenia*, „Zeszyty Naukowe” 2007, nr 2, s. 109–135.
- Searle J. R., *Minds, brains, and programs*, „The Behavioral and Brain Sciences” 1980, no. 3, s. 417–423.

Artificial intelligence as a substitute for the “religion” of modern man. Results of sociological research

Summary

The purpose of the publication was – by analyzing the results of sociological research – to show that artificial intelligence is a substitute for the new religion of modern man, understood in sociological terms. The article presents the results of a nationwide survey, conducted in 2023 on a sample of 1,052 people aged 15–65. A quantitative method, survey questionnaire technique (CAWI panel) was used. In the selection of the research sample—a random-quota method.

At the outset of the research process, the hypothesis was accepted that artificial intelligence can become a substitute for the religion of modern man. The hypothesis has been verified positively. Its confirmation is supported by the data that fit into the object-oriented and functional framing of the definition of religion.

Although AI becomes a substitute for an object of worship and performs some religious functions, it will never become a religion. It does not take into account the holistic concept of man, emphasizing the harmony of spirit and body. The problem of the soul, spirituality, is overlooked in the issue of artificial intelligence.

Keywords: artificial intelligence, new religion, substitute of religion, functions of religion

Sztuczna inteligencja i systemy ekspertowe w diagnozie i pomocy psychologicznej i psychoterapeutycznej

Wprowadzenie

Psychologia jest nauką, której głównym zadaniem jest zajmowanie się ludzkim doświadczeniem oraz tym, jak przekłada się ono na stany i cechy psychiczne. Mówimy zatem o doświadczeniu warunkującym zmienności wewnątrzsobnicze (stany) i międzysobnicze (cechy). Tym samym doświadczenie stanowi determinant, bodziec do przeżywania przez ludzi różnych stanów, które charakteryzuje wewnątrzsobnicza zmienność, gdyż nie są one zawsze stałe i takie same dla osoby – człowiek doświadcza różnych stanów. Zarazem doświadczenie stanowi także determinant, bodziec do kształtowania się cech psychicznych (na przykład osobowości), które charakteryzuje międzysobnicza zmienność. Oznacza to, że cechy te są bardziej stałe dla człowieka, za to ludzie różnią się między sobą w ich zakresie.

Psychologia zajmuje się zatem badaniem doświadczenia, stanów i cech psychicznych, ich zmiennością międzysobniczą i wewnątrzsobniczą.

¹ Dr hab. Agnieszka Szymańska, Instytut Psychologii UKSW.

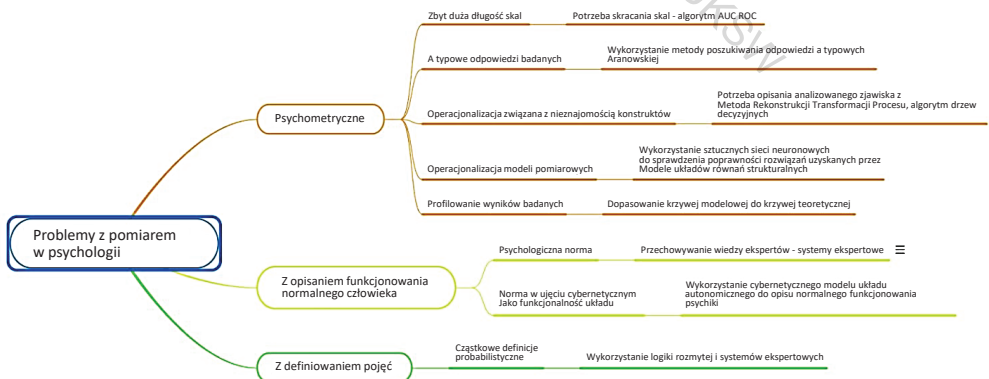
Ponieważ zarówno stany, cechy, jak i doświadczenia mają charakter cech niejawnych, ukrytych, niepodlegających bezpośrednio obserwacji, to stanowią illaty, cechy latentne. Narzuca to psychologii posługiwanie się odpowiednimi wskaźnikami pośrednimi, na podstawie których można inferować (wnioskować) o istnieniu cechy latentnej oraz jej nasileniu.

Oczywiście poszukiwanie tych wskaźników pozwalających na dobry pomiar cech ukrytych stanowi esencję całego pomiaru w psychologii. Chodzi bowiem o takie wyodrębnianie wskaźników, które będą zawierały w sobie wszystko to, co należy do badanego *indicatum*, a równocześnie odrzucały poza swój zakres wszystko, co do niego nie należy. Muszą się one równocześnie charakteryzować mocą korelacyjną, czyli występować wtedy, gdy zachodzi *indicatum*.

Poszukiwanie tych wskaźników stanowi duże wyzwanie. Ich odnalezienie wymaga spełnienia koniecznych warunków, aby powstała dobra metoda pomiaru psychologicznego. Dopiero ich spełnienie umożliwi stworzenie pomiaru, który ma szansę być rzetelnym i trafnym.

W tym tekście omówimy niektóre z tych koniecznych warunków dotyczących zarówno: a) pomiaru psychologicznego (psychometrii), jak i b) definicji pojęć oraz c) ogólnego problemu wynikającego z trudności opisu normalnego funkcjonowania człowieka. Wywód będziemy prowadzić zgodnie z mapą myśli przedstawioną na rysunku 1.

Rysunek 1. Mapa myśli dla omawianych w tekście zagadnień



Po przybliżeniu współczesnych potrzeb związanych z pomiarem w psychologii skupimy się na sposobach, w jaki sztuczna inteligencja może wyjść im naprzeciw. W jaki sposób może wesprzeć pomiar psychologiczny w zakresie do tej pory niewykorzystywanym w psychologii. Przybliżymy zatem ten problem w aspekcie wykorzystywania nie tylko algorytmów sztucznej inteligencji, ale także systemów ekspertowych oraz modeli cybernetycznych.

Należy pamiętać, że mówiąc o pomiarze w psychologii, odnosimy się do diagnozy psychologicznej i psychoterapeutycznej. Nie można bowiem odłączyć pomiaru od diagnozy, do której jest on wykorzystywany. Wzmacniając pomiar, wspieramy diagnostykę psychologiczną i psychoterapeutyczną. W tym tekście zostanie pokazane, jak proces diagnostyczny w pewnym stopniu już obecnie (a w przyszłości w zdecydowanie większym) może zostać wsparty przez sztuczną inteligencję.

Problemy z pomiarem w psychologii

Mówiąc o problemach z pomiarem z psychologii, musimy się odnieść do jego trzech aspektów: pierwszego, który dotyczy technicznych problemów pomiaru związanych z budową narzędzi i określaniem ich własności psychometrycznych; drugiego, który dotyczy przedmiotu tego pomiaru, a więc problemów związanych z definiowanymi pojęciami i wiedzą na temat mierzonych konstruktów; i trzeciego, który dotyczy interpretacji uzyskanych wyników i odnoszenia ich do normy, czyli do tego, co uznajemy w psychologii za „normalne”.

Wszystkie te aspekty sprowadzają się – rzecz jasna – do tego, co nazywamy trafnością i rzetelnością pomiaru psychologicznego². Wiadomo, że aby pomiar był trafny, wynik otrzymany w teście psychologicznym musi być zbliżony (tożsamy) z rzeczywistym stanem osoby badanej w zakresie mierzonej cechy. Odległość tych dwóch wyznacza błąd (*epsilon*). Przedstawia to poniższy wzór:

² A. Anastasi, S. Urbina, *Testy psychologiczne*, Pracownia Testów Psychologicznych Polskiego Towarzystwa Psychologicznego, Warszawa 1999.

$$O = P + \varepsilon$$

gdzie:

O – wynik otrzymany

P – wynik prawdziwy

ε – błąd pomiaru

Aby wynik otrzymany był jak najbardziej zbliżony do prawdziwego, test musi spełniać **konieczne warunki** pozwalające mu na stawianie trafnych diagnoz. Do tych warunków zalicza się: a) techniczne właściwości skal psychologicznych (psychometrycznych), b) definiowanie mierzonych konstruktów, c) odnoszenie wyników do normy, a więc aspekt samego stawiania diagnozy.

1.1. Techniczne aspekty skal psychologicznych

Jednym z kluczowych problemów współczesnych skal psychologicznych, obciążającym ich trafność, jest długość tych skal. Wiadomo, że długość procedury testowej ma wpływ na jakość wykonania zadań i – poza sytuacją, gdy procedura testowania dotyczy męczliwości – może mieć niekorzystny wpływ na trafność pomiaru. Szczególnie testy kliniczne, takie jak MMPI-2, mające ponad 500 pytań, budzą wątpliwości co do ich użyteczności diagnostycznej ze względu na swoją długość. Wiele profili jest nieinterpretowanych, gdyż zmęczeni badani mogą odpowiadać tendencyjnie lub przyspieszać odpowiedzi, co jest natychmiast wychwytywane przez skale kontrolne, i profil jest odrzucany. W przypadku długich narzędzi sama procedura testowania naraża badania na spadek trafności.

Jeszcze innym problemem są atypowe odpowiedzi badanych, którzy mają badaną cechę, a jednak ich odpowiedzi w testach są nietypowe, co sprawia, że diagnoza nie jest trafna. Takie atypowe odpowiedzi mogą się pojawiać z różnych powodów i dotyczyć tego, że cecha jest ujawniana – jak zauważa Aranowska – w sposób nietypowy przez niektóre osoby w populacji.

Również nieznanomość konstruktów, które podlegają pomiarowi, stanowi wielkie utrudnienie w konstrukcji skal psychologicznych. Badacz nie musi być nawet całkiem świadom, że wiedza na temat badanego konstruktu nie jest pełna. Powodów może być kilka, np. nie został on nigdy dokładnie

w literaturze przedmiotu opisany i/lub uległ z biegiem czasu przedawnieniu (szczególnie jeżeli zjawisko, które ma charakter historyczny, ulega zmianie w czasie)³.

Konieczność znajomości konstruktów jest *conditio sine qua non* poprawnego pomiaru psychologicznego. Ze złej znajomości konstruktów wynika bowiem problem trafności. Następuje wówczas niepoprawne zoperacjonalizowanie badanej cechy. Na poziomie modelu pomiarowego zmienne latentne będą wówczas bardziej związane z innymi zmiennymi niż ich zmienne obserwowalne z cechą latentną, do której należą. W takiej sytuacji model może mieć nawet bardzo dobre dopasowanie do danych, ale w rzeczywistości nie być w stanie niczego objaśnić⁴.

W końcu niektóre skale psychologiczne podają wynik nie w postaci wyniku ogólnego, ale w postaci profili. Dzieje się tak wówczas, gdy wyniki z poszczególnych skal nie dają się sprowadzić do wyniku ogólnego, gdyż skale nie są ze sobą skorelowane (nie tworzą metacechy). Wówczas każdą podskale traktuje się osobno, jako osobny wynik. Takie profile powstają np. w skalach osobowości, jak w skali Wielkiej Piątki, czy w skalach klinicznych, takich jak MMPI-2 lub CUIDA do badania rodziców adopcyjnych⁵.

Skąd jednak wiadomo, który profil jest tym poprawnym, a który nie? Zwykło się przyjmować, że poprawny to ten, dla którego poszczególne skale mieszczą się w konkretnych wynikach. Wiadomo jednak, że niekiedy pewne kształty profili wskazują na cechy wyższego rzędu, np. kombinacja cech Wielkiej Piątki, jaką stanowią: a) wysoki neurotyzm, b) niska sumienność, c) niska ugodowość, to profil charakterystyczny dla osobowości zaburzonych⁶.

Skąd jednak wiadomo, czy uzyskany przez konkretnego badanego profil jest bardziej podobny do zaburzonego czy nie? Nie znamy punktów

³ S. Nowak, *Metodologia badań społecznych*, PWN, Warszawa 2007.

⁴ A. Szymańska, *Predicting model for aggressive directiveness in the light of Tadeusz Tomaszewski's theory of action: structural and data mining approach*, „Psychology of Language and Communication” 2018, vol. 22(1), s. 354–371.

⁵ J. M. Digman, *Higher-order factors of the Big Five*, „Journal of Personality and Social Psychology” 1997, vol. 73, s. 1246–1256; T. Kucharski, J. Gomuła, *Wprowadzenie do kwestionariusza MMPI Wiskad*, [b.w.], Toruń 1998.

⁶ L. M. Saulsman, A. C. Page, *The five-factor model and personality disorder empirical literature: A meta-analytic review*, „Clinical Psychology Review” 2004, vol. 23(8), s. 1055–1085.

podziału tak potrzebnych do tego, aby móc określić bliskość profili, stąd często rozstrzygnięcia stawiane są „na oko”. W tym kontekście istotne jest opracowanie metod pozwalających na precyzyjne określenie profili występujących w populacjach oraz ich korelacji z innymi cechami. Dodatkowo konieczne jest zbadanie zgodności tych profili z założonymi teoretycznie krzywymi. W dalszej części artykułu zostanie pokazane, jak w tym celu można wykorzystać algorytmy.

1.2. Definiowanie pojęć psychologicznych

Nie wszystkie pojęcia wykorzystywane w psychologii mają zakres ostry, który niwelowałby problem z przypisaniem desygnatów do definiendum. Istnieją definicje probabilistyczne, w przypadku których wskazywane desygnaty pojęcia przynależą do niego z pewnym tylko prawdopodobieństwem. Pojęcia stanowiące definicje probabilistyczne mają charakter nieostry. Ich pomiar wymaga zastosowania zupełnie innej metody niż pomiar konstruktywów o charakterze ostrym. Trzeba bowiem w ich przypadku uwzględniać, że nie stwierdzamy o występowaniu cechy, stanu na podstawie wystąpienia wskaźnika. Musimy uwzględnić, że dany wskaźnik informuje o wystąpieniu definiendum z pewnym tylko prawdopodobieństwem lub przy założeniu istnienia jakichś dodatkowych przesłanek.

Innymi słowy definicje probabilistyczne zakładają istnienie pewnych warunków. Przykładem takiej definicji w psychologii jest pojęcie błędu wychowawczego Guryckiej. W definicji błędu wychowawczego autorka stwierdza, że jest to realne zachowanie rodzica, które związane jest z „ryzykiem” powstania negatywnych skutków dla rozwoju dziecka⁷. Negatywne skutki w tej definicji stanowią warunek konieczny, aby można było stwierdzić, że doszło do błędu wychowawczego. Do realnych zachowań rodzica, które mogą wiązać błąd wychowawczy przez wywołanie negatywnych dla rozwoju dziecka skutków, należą wyodrębnione w kole błędów wychowawczych zachowania uprawdopodobniające powstanie negatywnych skutków; są to: rygoryzm, agresja, hamowanie aktywności dziecka, obojętność,

⁷ A. Gurycka, *Błąd w wychowaniu*, WSiP, Warszawa 1990.

eksponowanie siebie przez rodzica, uleganie dziecku, zastępowanie dziecka w jego zadaniach, idealizacja dziecka⁸.

Zachowania opisane w kole błędów nie wiążą zatem błędu wychowawczego w sposób jednoznaczny, stanowią pewne jego (bardziej prawdopodobne niż inne zachowania rodzica) wyznaczniki. W takich sytuacjach istnieje niejaki prawdopodobieństwo, że dojdzie do powstania negatywnych skutków dla rozwoju dziecka.

Częstkowe definicje probabilistyczne są niezwykle użyteczne w naukach i równocześnie nastręczają sporo problemów z pomiarem. Wymagają zastosowania logiki rozmytej, którą trudno jest wykorzystać w testach psychologicznych ze względu na złożoność klucza. W dalszej części tekstu zostanie pokazane, jak w ich wykorzystaniu może pomóc sztuczna inteligencja i systemy ekspertowe.

1.3. Normalne funkcjonowanie człowieka

Psychologia do dziś nie dopracowała się również ustaleń na temat tego, czym jest **normalne** funkcjonowanie człowieka. Owszem, korzysta się w psychologii z pojęcia normy, ale oznacza ono porównanie wyniku badanego do odpowiadającej mu grupy odniesienia, i to w części odpowiadającej wynikom tej grupy. Normę zatem w psychologii wyznacza wynik w zakresie jakiejś cechy uzyskiwany w populacji. Nie mówi to nic o „funkcjonalności” osoby mającej dany wynik.

To niezwykle kontrowersyjne ujęcie normy. Gdyby bowiem większość populacji miała podwyższone wyniki w zakresie jakiejś cechy, np. depresji, wówczas zgodnie z ujęciem panującym obecnie w psychologii musielibyśmy za normę uznać stany depresyjne. Powiązanie normy z oczekiwanymi wartościami w populacji prowadzi do absurdów. Jednym z nich jest to, że norma ujęta w ten sposób nie ma właściwości rozwojowych, a wręcz blokuje rozwój teorii na temat poprawnego, właściwego funkcjonowania człowieka. Okazuje się bowiem, że nie trzeba opisywać normalnego funkcjonowania człowieka inaczej niż przez zastosowanie odniesienia do tego, jak zachowuje się

⁸ Tamże oraz A. Gurycka, *Błędy w wychowaniu*, w: *O wychowaniu i jego antynomiach*, red. E. Kubiak-Szyborska, D. Zajac, Wyd. WERS, Bydgoszcz 2008.

większość osobników w populacji. Właściwie jest to całkiem sensowne ujęcie, gdy chcemy porównywać ze sobą osoby – czemu w dużym stopniu służy pomiar w psychologii. Staje się jednak bezużyteczne, gdy mówimy o potrzebie opisanego normalnego zachowania człowieka. Pozbawieni jesteśmy wówczas wiedzy na temat tego, co jest „normalne”, funkcjonalne.

Odbiega zatem od normy bardzo wysoki wynik w skali wrażliwości (może on zostać uznany za zbyt sensorywny, dziecinny, znerwicowany), przynajmniej dopóki pozostałe osoby w populacji nie podniosą swoich wyników. Takie podejście wymusza zmianę na poziomie systemowym, ale ogranicza rozwój pojedynczych jednostek. Promuje zmiany generacyjne, ale nie indywidualne.

Gdyby psychologia знаła model normalnego funkcjonowania psychicznego i relacyjnego człowieka, wówczas rozwój poszczególnych osób nie byłby „uśredniony” do oczekiwanych wartości w populacji. Na podstawie wiedzy o cechach psychicznych psychologowie – wydaje się – nie będą w stanie wypracować takiego modelu. Nie wyznaczą bowiem odpowiedniego poziomu cechy – bo niby wobec czego, której innej cechy? W tym aspekcie psychologia nie umie sama wypracować takiego modelu. Zbudowanie modelu normalnego funkcjonowania byoby możliwe, jeżeli korzystano by z dorobku cybernetyki i jej wiedzy o sterowaniu układów samodzielnych⁹.

2. Wsparcie pomiaru psychologicznego za pomocą sztucznej inteligencji i systemów ekspertowych

Wraz z rozwojem nauki powstają nowe możliwości rozwiązywania problemów wcześniej – zdawałoby się – nierozwiązywalnych. Rozwój jednej dyscypliny naukowej wpływa na rozwój innej. Rozwój algorytmów sztucznej inteligencji w XX wieku miał wpływ na wiele dyscyplin: materiałoznawstwo, medycynę¹⁰,

⁹ M. Mazur, *Cybernetyczna teoria układów samodzielnych*, PWN, Warszawa 1966.

¹⁰ A. Sokołowski, J. Kosmol, *Diagnostyka narzędzia za pomocą sieci neuronowej*, Oficyna Wydawnicza Politechniki Wrocławskiej, 1995; E. Wołowicz-Korecka, *Zastosowanie sztucznych sieci neuronowych do modelowania procesów azotowania próżniowego stali narzędziowych*, „Przegląd

od jakiegoś czasu również na psychologię¹¹. Ich wdrażanie do psychologii może pomóc rozwiązać wiele zagadnień dotąd trudnych do rozwiązania.

Algorytmy sztucznej inteligencji są implementowane również w systemach ekspertowych, które dzięki ich zastosowaniu nie tylko są interaktywnymi bazami danych przechowującymi wiedzę ekspercką podaną im w postaci reguł przez człowieka, ale także, dzięki wspomnianym algorytmom, same te reguły odkrywają¹². Systemy ekspertowe mogą zatem stanowić coś na kształt „centrum badawczego”, w którym dochodzi do automatycznego przetwarzania (dzięki algorytmom) pozyskanych danych, a kolejno do zapisania reguł wynikających z obliczeń wykonanych przez algorytmy. Mogą rozwijać same siebie, korzystając z algorytmów, które pozwalają im się uczyć oraz adaptować do nowych danych, na przykład z internetu. Jednak samoistne rozwijanie się systemów ekspertowych wymaga nadzoru i interwencji ludzkiej.

W tym punkcie omówione zostaną możliwości, jakie stwarza sztuczna inteligencja i systemy ekspertowe, które są wykorzystywane w rozwiązywaniu problemów psychologii podniesionych w rozdziale pierwszym.

2.1. Sztuczna inteligencja a rozwój współczesnej psychometrii

Algorytmy sztucznej inteligencji mogą nie tylko w różny sposób wesprzeć diagnozę psychologiczną, ale też przyczynić się do rozwoju samej psychometrii. Dzięki ich zastosowaniu można wzmocnić pomiar, przyczynić się do jego większej rzetelności i trafności. Jak pokazała skonstruowana przez Ewę Rzechowską strategia Rekonstrukcji Transformacji Procesu, algorytm drzewa decyzyjnego C4.5 Quinlana może pomóc opisać dostarczone mu dane tak, że możliwe staje się odtworzenie ścieżek doświadczenia osób badanych. Tym samym analizowane zjawisko może podlegać opisowi w sposób ujmujący jego

Mechaniczny” 2020, nr 2, s. 25–31; StatSoftPolska 2016, s. 21–36, https://media.statsoft.pl/pdf/czytelnia/zastosowanie_sztucznych_sieci_neuronowych.pdf (3.12.2023).

¹¹ H. Y. B. Yim, Y. L. Boo, M. Ebbeck, *A Study of Children’s Musical Preference: A Data Mining Approach*, „Australian Journal of Teacher Education” 2014, vol. 39(2), s. 21–34.

¹² L. Rutkowski, *Metody i techniki sztucznej inteligencji*, PWN, Warszawa 2006.

transformację¹³. Ta niezwykle cenna strategia pozwala nie tylko opisywać nieznanе dotąd zjawiska – podobnie jak czyni to teoria ugruntowana – przewyższa tę ostatnią jednak w tym, że ostateczne poukładanie „ścieżek procesu” wykonuje nie człowiek, lecz algorytm sztucznej inteligencji. Algorytm C4.5 dokonuje takiego podziału elementów w zbiorze, aby wskazać różnice między nimi, wyodrębnić grupy elementów podobnych do siebie na podstawie informacji zawartych w zbiorze, do czego wykorzystuje indukcję matematyczną. Jest to więc jak widać matematyczna metoda rekonstrukcji zjawiska psychologicznego.

Jak pokazały Rzechowska i Szymańska, wykorzystanie tych atrybutów, które algorytm C4.5 użył do budowy drzewa decyzyjnego (czyli zmiennych wskazanych przez algorytm na drodze entropii), może skutkować zbudowaniem skali psychologicznej, do której dzięki algorytmowi zostały wykorzystane pytania niosące najwięcej informacji, a tym samym różnicujące najbardziej osoby badane w zakresie jakiejś cechy. Skalę taką będzie charakteryzowała nie tylko wysoka moc dyskryminacyjna pytań, ale również trafność, a tym samym rzetelność¹⁴.

Algorytmy mogą być wykorzystane w psychologii również do sprawdzenia poprawności rozwiązań otrzymanych w procedurze modelowania za pomocą układów równań strukturalnych. Jak pokazały Aranowska i Szymańska, model strukturalny może być dopasowany do danych, a jednak mieć słabą predykcję¹⁵. Wynika to z niepoprawnie skonstruowanego modelu pomiarowego i dopuszczenia modelu, w którym zmienne latentne lepiej objaśniają siebie nawzajem niż są objaśniane przez własne zmienne obserwowalne (istnieją wyższe związki na ścieżkach między zmiennymi latentnymi niż ładunki czynnikowe lambda dla zmiennych obserwowalnych). Innymi słowy, zaburzona jest moc dyskryminacyjna w takim modelu, o którym

¹³ E. Rzechowska, *Potencjalność w procesie rozwoju: Mikroanaliza konstruowania wiedzy w dziecięcych interakcjach rówieśniczych*, Wydawnictwo KUL, Lublin 2004.

¹⁴ E. Rzechowska, A. Szymańska, *Wykorzystanie strategii Rekonstrukcji Transformacji Procesu do budowy skali psychologicznej*, w: *Diagnozowanie – wyzwania i konteksty*, red. W. J. Paluchowski, Wydawnictwo Naukowe UAM, Poznań 2017, s. 31–58.

¹⁵ E. Aranowska, A. Szymańska, *Trafność zmiennej latentnej wyjaśnianej przez model SEM*. Referat podczas Ogólnopolskiej Konferencji Naukowej „Katowickie Spotkania Psychometryczne”, Katowice 2017.

pisze się w literaturze przedmiotu¹⁶. W przypadku takich modeli ich trafność jest niska, jak wynika ze wzorów Aranowskiej¹⁷. Rozwiązanie otrzymane ze wzoru Aranowskiej potwierdzają również sztuczne sieci neuronowe; mogą one zostać wykorzystane do sprawdzenia, czy model jest trafny¹⁸. Innymi słowy: czy na podstawie zmiennych w modelu da się sensownie przewidywać wyniki głównej zmiennej objaśnianej przez model. Takie rozwiązania były już w psychologii wykorzystywane¹⁹.

Algorytmy matematyczne mogą wesprzeć również proces skracania skal i wskazywania najlepszych pozycji testowych. Jak pokazał Koczkodaj i zespół, wykorzystanie algorytmu AUC ROC przy wiedzy na temat zewnętrznego kryterium diagnostycznego umożliwiło wyznaczenie w skali pozycji testowych, które miały największą moc diagnostyczną. To umożliwiło wyłonienie najlepszych pozycji i zredukowanie liczby pozycji w skali aż o dwie trzecie. Nie pogorszyło to równocześnie parametrów psychometrycznych skali²⁰.

Aplikacja algorytmów *data mining*, wykorzystujących analizę skupień k-średnich, umożliwia natomiast wykreślenie krzywej empirycznej (pochodzącej z populacji), co pozwala na określenie jej dopasowania do krzywej modelowej (wykreślonej na poziomie teoretycznym). Dzięki wykorzystaniu współczynników Aranowskiej można określać dopasowanie obu krzywych do siebie²¹. Logika tej procedury podobna jest do modelowania wykorzystywanego w modelach strukturalnych SEM. Za pomocą współczynników szacuje się, na ile krzywa teoretyczna jest dopasowana do krzywej empirycznej. Podsumowując: wykorzystanie algorytmów sztucznej inteligencji oraz innych algorytmów służących do odkrywania wzorców w danych może znacznie wesprzeć proces diagnozy oraz konstrukcji skal psychologicznych.

¹⁶ J. J. Hair, W. C. Black, B. J. Babin, R. E. Anderson, R. L. Tatham, *Multivariate data analysis*, Upper Saddle River, NJ: Pearson 2006.

¹⁷ E. Aranowska, A. Szymańska, *Trafność zmiennej latentnej...*, dz. cyt.

¹⁸ Tamże.

¹⁹ A. Szymańska, *Predicting model...*, art. cyt.

²⁰ W. W. Koczkodaj, T. Kakiashvili, A. Szymańska, J. Montero-Marin, R. Araya, J. Garcia-Campayo, K. Rutkowski, D. Strzałka, *How to reduce the number of rating scale items without predictability loss?*, „Scientometrics” 2017, 111(2), s. 581–593.

²¹ E. Aranowska, *Wskaźniki bliskości dowolnej krzywej do krzywej modelowej*, „Psychologia Matematyczna” 1989, nr 3, s. 99–114.

2.2. Operacjonalizacja pojęć probabilistycznych oraz diagnoza psychologiczna w systemach ekspertowych

Współczesne pojęcia psychologiczne operują definicjami probabilistycznymi. Do ich operacjonalizacji potrzeba wykorzystać logikę rozmytą, gdyż nie można jednoznacznie określić przynależności desygnatów do ich pojęcia; można to zrobić z pewnym tylko prawdopodobieństwem. Tak operacjonalizację pojęcia błędów wychowawczych pokazała Szymańska, operacjonalizując reakcję dziecka na sytuację wychowawczą²². Wykorzystana do tego logika rozmyta umożliwiła przypisanie różnym odczuciom przynależności do zbioru uczuć pozytywnych lub negatywnych. Kolejno, w podobny sposób, były klasyfikowane zachowania dziecka, to, co pomyślało, informacje, jakie wyciągnęło z sytuacji, reaktancja dziecka, poziom zaburzenia relacji z rodzicem. Po dokonaniu tych klasyfikacji – określenie, z jakim prawdopodobieństwem doszło do błędu wychowawczego na podstawie doświadczenia dziecka²³.

Takie opisanie doświadczenia dziecka jest bardzo czasochłonne, a ewentualny klucz do skali, która bada doświadczenie dziecka, tak bardzo złożony, że właściwie od początku trzeba myśleć o umieszczeniu skali w systemie ekspertowym. Będzie on nie tylko przeprowadzał badania, ale także obliczał wyniki i udzielał informacji zwrotnej.

System ekspertowy nie tylko będzie dysponował wiedzą na temat klucza do skali, ale może też zostać wzbogacony o dodatkową wiedzę ekspercką z zakresu błędów wychowawczych, którą będzie mógł brać pod uwagę podczas stawiania diagnozy. Co więcej, szkielety systemów ekspertowych mają zaimplementowane algorytmy, takie jak drzewa decyzyjne czy sztuczne sieci neuronowe²⁴. Mają więc możliwości przetwarzania danych, które są im dostarczane, i tworzenia nowych reguł, dotąd nieznanych.

²² A. Szymańska, *Predicting model...*, art. cyt.

²³ A. Szymańska, *Strategia Badania Doświadczenia Dzieci na przykładzie sytuacji hamowania ich aktywności*, „Studia z Teorii Wychowania” 2019, vol. 1(26), s. 193–216.

²⁴ K. Michalik, *Neuronix. Symulator sztucznych sieci neuronowych*, AITECH, Katowice 2006. Tenże, *Systemy ekspertowe we wspomaganii procesów zarządzania wiedzą w organizacji*, Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, Katowice 2014, s. 186–193: rozdz. „HybRex jako narzędzie do budowy hybrydowych systemów wspomaganii decyzji”.

Normalne funkcjonowanie człowieka w ujęciu cybernetycznym

Opisanie, czym jest normalne funkcjonowanie, jawi się jako nie lada sztuka, której psychologia do tej pory się nie dopracowała. W tym celu może jednak skorzystać z pomocy innych dyscyplin naukowych, a szczególnie tej, która zajmuje się opisywaniem funkcjonowania wszystkich maszyn samosterownych, autonomicznych (w tym człowieka), czyli cybernetyki. Gdyby wykorzystać wiedzę z cybernetyki do opisu normalnego funkcjonowania człowieka, to przede wszystkim funkcjonowanie takie charakteryzowałyby dwie cechy: a) umiejętność zachowania wewnętrznej równowagi (homeostazy) oraz b) funkcjonalność, czyli skuteczność funkcjonowania umożliwiająca człowiekowi wywieranie wpływu na swoje otoczenie²⁵.

A zatem człowiek normalny to ten, który potrafi tak równoważyć swoje procesy psychiczne, aby skutecznie wywierać wpływ na własne otoczenie. Opisując funkcjonowanie osobowości depresyjnej w ujęciu psychocybernetycznym, Szymańska pokazała, jak opisać funkcjonowanie psychiczne człowieka za pomocą modelu cybernetycznego systemu autonomicznego²⁶. Sprowadza się ono do równowagi przepływu potencjałów, nie dopuszczając ani do reakcji zbyt impulsywnych, ani do całkowitego braku reakcji. Jest to takie funkcjonowanie, które nie dopuszcza również do odkładania się energii w ciele, co mogłoby doprowadzić do bloków mięśniowych, a w konsekwencji do zaburzeń psychicznych²⁷.

Na czym polega wywieranie wpływu na otoczenie? W ujęciu cybernetycznym to takie zachowanie, które umożliwia człowiekowi pozyskiwanie energii koniecznych do dwóch celów: a) przetrwania oraz b) dalszego rozwijania się. Z tym że w przypadku dalszego rozwoju osoba normalna musi zachować równowagę. Zbyt szybki rozwój sprawia bowiem, że system się szybciej starzeje, a to skraca mu żywotność. Zatem nawet rozwój musi być zrównoważony. Model cybernetyczny ukazuje, że zbyt szybki rozwój wymaga od systemu pozyskiwania energii na utrzymanie swojego procesu,

²⁵ M. Mazur, *Cybernetyka i charakter*, PIW, Warszawa 1976.

²⁶ Tamże.

²⁷ A. Lowen, *Wstęp do bioenergetyki*, Jacek Santorski Co Agencja Wydawnicza, Warszawa 1992.

i to może skrócić jego żywotność. Model cybernetyczny ukazuje, że nawet rozwój musi być zrównoważony, inaczej bowiem przestaje być opłacalny. Zatem człowiek normalny, poprawnie funkcjonujący, dąży do rozwoju, ale bez przesady. Robi to w sposób równomierny, zrównoważony. Nie pozwala sobie ponosić zbyt dużych kosztów z tego tytułu.

Niezwykle ciekawe wnioski z modelu cybernetycznego wypływają również dla podatności, która charakteryzuje osobę normalną. Podatność to inaczej umiejętność znoszenia sytuacji, które nie pasują do naturalnej dynamiki osoby. Każdy człowiek ma pewne sytuacje, które lubi i mu pasują, ale też takie, które mu nie pasują. Jedna osoba np. lubi uległość i dobrze się czuje w sytuacjach, w których musi podporządkowywać się innym, inna znowu woli sytuacje dominacji, gdy to ona może dominować. Na przykład osoba, która lubi dominację, dobrze znosi sytuacje wymagające od niej bycia dowódcą, dominowania; równocześnie potrafi tolerować sytuacje, w których nie ma władzy i nie dominuje, ale może już z trudem przeżywać sytuacje wymagające od niej uległości i podporządkowania. Z kolei osoba uległa będzie wolała sytuacje wymagające od niej uległości, ale źle znosić te, w których to ona musi dominować. Takie sytuacje mogą okazać się dla niej całkowicie nieakceptowalne.

Nie ma ludzi całkowicie podatnych, to znaczy takich, którzy są w stanie przyjąć każdą sytuację i w każdej sytuacji funkcjonować, ale wydaje się, że osoba normalna skutecznie funkcjonująca potrafi sobie poradzić w różnych sytuacjach. Ma dość szeroką podatność na różne sytuacje, umie się w nich odnaleźć. W przeciwieństwie do niej osoby o zaburzonej osobowości mają znacznie węższą podatność i dotyczy ona pewnych skrajnych zachowań: dominacji albo totalnej uległości. A zatem te osoby potrafią głównie dominować i nie znoszą uległości, przestrzegania zasad (tak jak osoby antyspołeczne) lub potrafią być zależne i uległe, a zupełnie nie potrafią znaleźć się w sytuacjach, gdy mają władzę i mogą dominować (np. osobowości zależne)²⁸.

Oczywiście model cybernetyczny umożliwi opisanie procesów zachodzących w osobowości zdrowej w dużo dokładniejszy sposób. W tym miejscu wymienia się jedynie niektóre, i to dość ogólne cechy. Ważne jest natomiast,

²⁸ T. Millon, R. Davis, *Disorders of Personality: DSM-IV and Beyond* (2nd ed.), John Wiley and Sons, Oxford 1996.

że model cybernetyczny umożliwia takie opisanie osoby normalnej, aby nie tylko dość wyczerpująco ją scharakteryzować, ale również umożliwić odróżnianie od osób, których funkcjonowanie odbiega od normy, a więc – rozumując w języku cybernetycznym – jest niefunkcjonalne.

Dalszy rozwój sztucznej inteligencji dzięki naukom społecznym

Rozwój sztucznej inteligencji (AI), dzięki współpracy z naukami społecznymi, otwiera nowe perspektywy na przyszłość technologii i społeczeństwa. Jak już wykazano, postępy w AI mogą znacząco wspierać psychologię, zarówno w zakresie pomiaru psychologicznego, jak i opisu normalnych procesów psychicznych. Istnieje jednak pytanie o sytuację odwrotną: Jak psychologia może przyczynić się do rozwoju AI, szczególnie w kontekście definiowania „normy” funkcjonowania?

W obliczu rosnących obaw przed potencjalnymi zagrożeniami ze strony AI, takimi jak dominacja nad ludzkością czy narzucanie sposobu myślenia, ważne jest zrozumienie, że te lęki wynikają głównie z jej potężnych możliwości obliczeniowych. Dzięki algorytmom indukcyjnym i na podstawie dostępnych danych technologia AI ma zdolność generowania nowych reguł, co potęguje jej możliwości poznawcze.

Obawa, że przerośnie ludzi pod względem wiedzy, to tylko jeden z niepokojów. Boimy się, że będąc pozbawioną uczuć jako „czysta inteligencja”, może stać się psychopatyczna, że będzie przypominała inteligentnego, dążącego do celu psychopatę, który nie będzie taki jak my pod względem uczuciowym. Nie akceptujemy psychopatii w życiu społecznym wśród ludzi, a co dopiero myśl, że stworzymy sztucznego psychopatę! Nic dziwnego, że podnoszone są obawy co do sztucznej inteligencji.

Jednak to nie uczuciowość, lecz wspólne wartości i normy etyczne stanowią o naszej „ludzkości”. W tej perspektywie możemy „nauczyć” AI odczuwania i przestrzegania tych wartości, nadając jej algorytmom odpowiednie wagi. Ważne jest jednak, by pamiętać, że takie „odczuwanie” przez AI nie czyni jej identyczną z ludźmi – równowaga emocjonalna i homeostaza

są cechami charakterystycznymi dla zdrowego funkcjonowania, a nie wyłącznie dla ludzi.

Pojęcie „czucia” w kontekście AI należy interpretować jako zdolność do symulowania reakcji emocjonalnych, a nie dosłowne doświadczanie uczuć. Systemy AI mogą być programowane do rozpoznawania i reagowania na sygnały emocjonalne w sposób naśladujący ludzkie reakcje, co ma na celu poprawę interakcji z użytkownikami i dostosowanie do ich potrzeb emocjonalnych. Jest to jednak wynik zaawansowanych algorytmów i analizy danych, a nie dowód na posiadanie przez maszyny wewnętrznego życia emocjonalnego. Takie wyjaśnienie pomaga rozwiązać wątpliwości dotyczące antropomorfizowania AI i podkreśla jej narzędziowy charakter.

Wyzwaniem jest zatem określenie, co stanowi „normalne” funkcjonowanie AI i jak możemy to osiągnąć przez programowanie. Bez jasno zdefiniowanej „normy” pozostajemy z ograniczonym potencjałem technologii i koniecznością regulacji jej działania prawnie, co na dłuższą metę może okazać się nieskuteczne. „Uspołecznienie” AI, czyli wprowadzanie jej w ramy społeczne i etyczne, wydaje się najbardziej obiecującym sposobem na kontrolę jej rozwoju.

W procesie tzw. wychowania AI warto skupić się na aspekcie programowania etycznego i wprowadzania norm społecznych, co pozwala na lepsze zrozumienie roli sztucznej inteligencji jako narzędzia służącego człowiekowi. Tym samym termin „wychowanie” nie odnosi się do procesu socjalizacji analogicznie do ludzkiego, lecz do starannego projektowania algorytmów i systemów uczących się, które działają w ramach ściśle określonych przez twórców zasad etycznych. Takie podejście podkreśla, że sztuczna inteligencja nie jest autonomicznym bytem zdolnym do samodzielnego rozwoju moralnego, ale narzędziem kształtowanym przez ludzi, mającym działać zgodnie z wartościami, które uznajemy za pożądane.

Cybernetyka i inne dziedziny naukowe mogą odgrywać kluczową rolę w projektowaniu AI, która nie tylko „czuje” i podejmuje decyzje, ale również jest zdolna do utrzymania własnej emocjonalnej równowagi. To otwiera przed AI możliwości, by stała się dla ludzkości partnerem, a nie jej zagrożeniem.

Wnioski dotyczące możliwości i ograniczeń sztucznej inteligencji powinny być formułowane z uwagą na obecną wiedzę naukową i technologiczną. Ważne jest oddzielenie faktów i osiągnięć obecnych technologii

od spekulacji na temat przyszłego rozwoju AI. Uściślenie tego, co AI może robić teraz i co potencjalnie będzie w stanie robić w przyszłości, pomaga w uniknięciu nadinterpretacji jej zdolności. Równocześnie należy podkreślić, że wszelkie dyskusje o „uczuciach” AI i jej „wychowaniu” służą przede wszystkim zrozumieniu i zarządzaniu potencjalnymi ryzykami związanymi z jej rozwojem.

Podsumowanie

Rozwój sztucznej inteligencji (AI) w kontekście nauk społecznych otwiera przed nami fascynujące możliwości udoskonalenia zarówno pomiaru psychologicznego, jak i głębszego zrozumienia normalnego funkcjonowania psychicznego. Współpraca między tymi dyscyplinami sugeruje, że potencjalne korzyści dla rozwoju AI, wynikające z wdrożenia przy jej projektowaniu modeli psychologicznych, mogą być szerokie i dalekosiężne, choć w obecnym momencie trudne do precyzyjnego przewidzenia. Skutkiem bezpośrednim takiej współpracy jest głębsze zrozumienie norm społecznych i etycznych, które mogą kształtować rozwój technologiczny.

Mimo że dyskusje na temat zagrożeń płynących ze strony AI są powszechne – od obaw o dominację nad ludzkością po zmianę sposobu myślenia i życia – warto zauważyć, że te obawy wynikają głównie z jej rosnących możliwości obliczeniowych. Technologia AI, dzięki zaawansowanym algorytmom, jest w stanie generować nowe reguły i wnioski, co nieuchronnie rozszerza granice wiedzy.

Jednakże rozważając kwestię „uczuciowości” i decyzji moralnych AI, musimy być precyzyjni w naszym języku. „Wychowanie” AI nie oznacza nadawania jej ludzkich cech czy świadomości, lecz projektowanie systemów, które działają w zgodzie z wyznaczonymi normami etycznymi i społecznymi. Tym samym, kiedy mówimy o „uczuciach” AI, odnosimy się do zdolności symulowania reakcji emocjonalnych w interakcjach z ludźmi, a nie do dosłownego doświadczania emocji. Jest to ważne wyjaśnienie, które pomaga rozwiązać wszelkie wątpliwości dotyczące antropomorfizowania AI.

Przyjmując powyższe przesłanki, należy podkreślić, że wszelkie nawiązania do szacunku dla „istnienia” AI odnoszą się do odpowiedzialnego

wykorzystywania tej technologii, z uwzględnieniem jej wpływu na społeczeństwo, a nie do przypisywania jej statusu moralnego podmiotu. W tym kontekście „uspołecznienie” AI, czyli wprowadzanie jej w kontekst norm społecznych i etycznych, staje się kluczowe dla bezpiecznego i odpowiedzialnego rozwijania technologii.

Rozwój AI, uzupełniony o wiedzę z zakresu nauk społecznych, ma potencjał przyczynienia się nie tylko do wielkich odkryć w psychologii, ale również do znacznego ulepszenia pomocy psychologicznej i interwencji psychoterapeutycznych. Dzięki głębszemu zrozumieniu normalnego funkcjonowania człowieka możliwe staje się projektowanie systemów AI, które mogą wspierać rozwój psychologiczny i terapeutyczny na niespotykaną dotąd skalę.

Podsumowując, należy zaznaczyć, że wszelkie dyskusje o „uczuciach” AI i jej „wychowaniu” mają na celu nie antropomorfizację, lecz zrozumienie i zarządzanie potencjalnymi ryzykami związanymi z jej rozwojem. Rozważania te powinny być prowadzone z należytą starannością i z uwzględnieniem obecnej wiedzy naukowej i technologicznej, a także etycznych i społecznych implikacji stosowania AI. Ostatecznie odpowiedzialny rozwój AI, który respektuje zarówno możliwości technologiczne, jak i ludzkie wartości, jest kluczowy dla harmonijnego współistnienia technologii i społeczeństwa.

Bibliografia

- Anastasi A., Urbina S., *Testy psychologiczne*, Pracownia Testów Psychologicznych Polskiego Towarzystwa Psychologicznego, Warszawa 1999.
- Aranowska E., *Wskaźniki bliskości dowolnej krzywej do krzywej modelowej*, „Psychologia Matematyczna” 1989, nr 3, s. 99–114.
- Aranowska E., Szymańska A., *Trafność zmiennej latentnej wyjaśnianej przez model SEM*, Referat podczas Ogólnopolskiej Konferencji Naukowej „Katowickie Spotkania Psychometryczne”, Katowice 2017.
- Digman J. M., *Higher-order factors of the Big Five*, „Journal of Personality and Social Psychology” 1997, vol. 73, s. 1246–1256.
- Gurycka A., *Błąd w wychowaniu*, WSiP, Warszawa 1990.
- Gurycka A., *Błędy w wychowaniu*, w: *O wychowaniu i jego antynomiach*, red. E. Kubiak-Szyborska, D. Zając, Wyd. WERS, Bydgoszcz 2008.
- Hair J. J., Black W. C., Babin B. J., Anderson R. E., Tatham R. L., *Multivariate data analysis*, Upper Saddle River, NJ: Pearson 2006.

- H. Y. B. Yim, Y. L. Boo, M. Ebbeck., *A Study of Children's Musical Preference: A Data Mining Approach*, „Australian Journal of Teacher Education” 2014, vol. 39(2), s. 21–34.
- Koczkodaj W. W., Kakiashvili T., Szymańska A., Montero-Marin J., Araya R., Garcia-Campayo J., Rutkowski K., Strzałka D., *How to reduce the number of rating scale items without predictability loss?*, „Scientometrics” 2017, 111(2), s. 581–593.
- Kucharski T., Gomuła J., *Wprowadzenie do kwestionariusza MMPI Wiskad*, [b.w.], Toruń 1998.
- Lowen A., *Wstęp do bioenergetyki*, Jacek Santorski & Co Agencja Wydawnicza, Warszawa 1992.
- Mazur M., *Cybernetyczna teoria układów samodzielnych*, PWN, Warszawa 1966.
- Mazur M., *Cybernetyka i charakter*, PIW, Warszawa 1976.
- Michalik K., *Neuronix. Symulator sztucznych sieci neuronowych*, AITECH, Katowice 2006.
- Michalik K., *Systemy ekspertowe we wspomaganiu procesów zarządzania wiedzą w organizacji*, Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, 2014, s. 186–193: rozdz. „HybRex jako narzędzie do budowy hybrydowych systemów wspomaganie decyzji”.
- Millon T., Davis R., *Disorders of Personality: DSM-IV and Beyond* (2nd ed.), John Wiley and Sons, Oxford 1996.
- Nowak S., *Metodologia badań społecznych*, PWN, Warszawa 2007.
- Rutkowski L., *Metody i techniki sztucznej inteligencji*, PWN, Warszawa 2006.
- Rzechowska E., *Potencjalność w procesie rozwoju. Mikroanaliza konstruowania wiedzy w dziecięcych interakcjach rówieśniczych*, Wyd. KUL, Lublin 2004.
- Rzechowska E., Szymańska A., *Wykorzystanie strategii Rekonstrukcji Transformacji Procesu do budowy skali psychologicznej*, w: *Diagnozowanie – wyzwania i konteksty*, red. W. J. Paluchowski, Wydawnictwo Naukowe UAM, Poznań 2017, s. 31–58.
- Saulsman L. M., Page A. C., *The five-factor model and personality disorder empirical literature: A meta-analytic review*, „Clinical Psychology Review” 2004, vol. 23(8), s. 1055–1085.
- Sokołowski A., Kosmol J., *Diagnostyka narzędzia za pomocą sieci neuronowej*, Oficyna Wydawnicza Politechniki Wrocławskiej, 1995.
- Szymańska A., *Predicting model for aggressive directiveness in the light of Tadeusz Tomaszewski's theory of action: structural and data mining approach*, „Psychology of Language and Communication” 2018, vol. 22(1), s. 354–371.
- Szymańska A., *Strategia Badania Doświadczenia Dzieci na przykładzie sytuacji hamowania ich aktywności*, „Studia z Teorii Wychowania” 2019, vol. 1(26), s. 193–216.
- Wołowicz-Korecka E., *Zastosowanie sztucznych sieci neuronowych do modelowania procesów azotowania próżniowego stali narzędziowych*, „Przegląd Mechaniczny” 2020, nr 2, s. 25–31; StatSoftPolska 2016, s. 21–36, https://media.statsoft.pl/pdf/czytelnia/zastosowanie_sztucznych_sieci_neuronowych.pdf (3.12.2023).

Artificial intelligence and expert systems in psychological and psychotherapeutic diagnosis and assistance

Summary

The text discusses the problem of using artificial intelligence algorithms to improve: 1) psychological measurement and in particular: a) construction of new scales, b) shortening the length of psychological scales, c) validation of structural models, d) plotting profiles of subjects;

2) operationalize probabilistic definitions using fuzzy logic and software them in expert systems;

3) using the cybernetic model of the autonomous system to describe the psychological norm.

Possible benefits for the development of artificial intelligence are also presented, which can be consequences of the development of disciplines such as sociology, pedagogy, philosophy, psychology. It seems that they can contribute to the development of artificial intelligence by describing in cybernetic models how the psychological norm functions, and correct social functioning. Such a description could be used to develop artificial intelligence programs. It could make it more similar to humans not only in terms of their thinking (as artificial intelligence algorithms were designed to do) but also in terms of their experiencing feelings (even “artificial feelings”) and values, social norms.

Keywords: Measurement, psychometrics, artificial intelligence, expert system, cybernetic

Rola sztucznej inteligencji w walce z mową nienawiści w mediach społecznościowych na przykładzie Facebooka i Twittera

W mediach społecznościowych (*social media*) dokonuje się masowe wysyłanie treści, które budzą oburzenie i kontrowersje, prowokując platformy do reakcji. Wykorzystywanie internetu tak do polaryzacji społeczeństwa, jak i wojny hybrydowej między Rosją a Ukrainą przynosi nowe wyzwania związane z mową nienawiści. Duże nadzieje pokładane są w systemach uczenia maszynowego, które zrobią to szybciej, sprawniej, a zdaniem niektórych – skuteczniej niż człowiek. Zjawisko sztucznej inteligencji, powstałe na styku informatyki, robotyki i matematyki, ma dzisiaj charakter interdyscyplinarny, jest przedmiotem zainteresowania nauk społecznych, nauki o zdrowiu i praktycznych dziedzin życia, takich jak transport czy marketing. Będzie wyzwaniem dla środowiska gospodarczego, usług publicznych i, jak się okazuje, także dla mediów i technologii oferujących komunikację przez internet.

¹ Dr Anna Leszczuk-Fiedziukiewicz, Wydział Socjologii, Uniwersytet w Białymstoku.

Cele, założenia i metody badawcze

Autorka stara się odpowiedzieć na pytania o szanse i zagrożenia stojące przed stosowaniem przez Facebook i Twitter sztucznej inteligencji w moderowaniu treści cechujących się ksenofobią, rasizmem, upokarzaniem i pogardzaniem, które w polskim dyskursie socjologicznym określane są mianem mowy nienawiści. Inspiracją do zajęcia się tym tematem jest dyskusja w przestrzeni publicznej o fali hejtu w internecie i oczekiwanie na reakcję właścicieli mediów społecznościowych. Ich aktywność na tym polu cechuje biegunowość, ponieważ deklaracje o angażowaniu się w działania poprawiające bezpieczeństwo internautów i tzw. czyszczenie internetu stoją obok unikania odpowiedzialności za ingerowanie w treści w obawie przed zarzutami o uprawianie cenzury, mającego dużo wspólnego z postawą nonszalancji wobec etycznych rozwiązań. Dodatkowym efektem niniejszego artykułu jest jego funkcja edukacyjna, mająca na celu przybliżenie społecznego aspektu sztucznej inteligencji (AI, SI), ponieważ zgodnie z badaniami w percepcji Polaków jest ona kojarzona z wieloma zjawiskami. Jeszcze w 2019 r. najczęstszymi wyobrażeniami związanymi z AI były: humanoidalny robot (mówiący i poruszający się tak jak człowiek; 69%), bardzo zaawansowany komputer (54%) lub autonomiczny samochód (30%)².

Metody badawcze wywodzą się z nauk społecznych i obejmują analizę danych zastanych, czyli publikacji dotyczących aktywności Facebooka i Twittera oraz deklaracji i decyzji biznesowych tych firm. Inne metody wykorzystane w publikacji to analiza dyskursu medialnego związanego z wykorzystaniem sztucznej inteligencji przez wspomniane platformy internetowe. Dotyczyły one informacji z portali internetowych zajmujących się komunikacją i rozwiązaniami technologicznymi stosowanymi przez social media. Do analizy wybrano wypowiedzi właścicieli firm oraz komentarze ekspertów interpretujących deklaracje i postulaty Marka Zuckerberga i Elona Muska.

Jedną z metod badania była analiza treści publikacji dotyczących wizerunku Facebooka, a zwłaszcza jego roli w organizowaniu globalnej

² M. Nowina-Konopka, *Każdy zna się na AI. Przegląd badań polskiej opinii publicznej na temat sztucznej inteligencji*, „Zeszyty Prasoznawcze” 2023, t. 66, nr 4 (256), s. 132.

przestrzeni komunikacyjnej. Spinającą je klamrą była ewolucja polityki platformy, chociaż publikacje były różnorodne pod względem gatunkowym (reportaż, książka popularnonaukowa, praca naukowa, poradnik). Książka *Nabici w Facebooka. Przestroga przed katastrofą* (wyd. oryg. 2019), oprócz tytułu intryguje tym, że napisał ją doradca Marka Zuckerberga, Roger McNamee – początkowo inwestor, potem krytyk firmy przestrzegający przed jej manipulacjami. Sheera Frenkel i Cecilia Kang wydały książkę *Brzydka prawda. Kulisy walki Facebooka o dominację* (wyd. oryg. 2021) o przyczynach procesów sądowych wytoczonych firmie Zuckerberga. Śledztwo prowadzone przez dziennikarki „New York Timesa” polegało na nieformalnym dojściu do dokumentów Facebooka i rozmowach z jego pracownikami. Publikacja *A miało być tak pięknie. Za kulisami imperium z Doliny Krzemowej* (wyd. oryg. 2020) Stevena Levy’ego to efekt wieloletniej pracy, polegającej na analizie dyskursu prasowego nt. Facebooka, i przeprowadzenia wywiadów z pracownikami firmy oraz ekspertami znającymi świat technologii i mediów społecznościowych. Celem autora było pokazanie nieetycznych decyzji skutkujących zmianą wizerunku firmy z medium, które łączy ludzi, w technologię, która ich skłóca. Wszystkie te publikacje wiąże ponadto obnażenie działań technologicznego giganta, który cały czas eksperymentuje ze swoimi algorytmami. W książce *Antisocial Media. Jak Facebook oddala nas od siebie i zagraża demokracji* (wyd. oryg. 2018) Siva Vaidhyanathan, profesor Uniwersytetu Virginia, medioznawca i historyk kultury, bezlitośnie analizuje działania zespołu Marka Zuckerberga. *Władza platform. Za fasadą Google, Facebooka i Spotify* (2022) Jana Krefta stanowi z kolei naukowe podejście do polityki technologicznych gigantów i jej konsekwencji w wymiarze społecznym. Deklaracje szefów Facebooka i Twittera oraz komentarze eksperckie pochodziły z portali AI Business, CyberDefence24, Wszystko co Najważniejsze, Wirtualne Media, Fronda i kilku redakcji anglojęzycznych („New York Post”, BBC, Reuters).

Definicje mediów społecznościowych, sztucznej inteligencji i mowy nienawiści

Media społecznościowe rozumiane są jako grupy aplikacji opartych na internecie, które zostały stworzone na ideologicznym i technologicznym fundamencie Web 2.0 i które umożliwiają tworzenie i wymianę treści generowanych przez użytkowników³. Facebook jest monopolistą usług on-line, kontrolującym ponad 70% ruchu w sieci, nazywany przez M. Jużę „bramą, przez którą musi przejść każdy użytkownik Internetu”⁴, ponieważ jest też właścicielem innych serwisów: Instagrama, Messengera czy WhatsAppa. Pełni wiele społecznych funkcji nie tylko wśród młodego pokolenia, jest innowacyjnym miejscem o funkcji tożsamościowej i komunikacyjnej, pomaga w poszukiwaniu informacji, wyrażaniu siebie i podtrzymywaniu relacji⁵. Facebook jako jeden z pierwszych start-upów z Doliny Krzemowej, dokładnie od 2006 r., nagłośnił używanie sztucznej inteligencji w działaniach reklamowych i doświadczeniu użytkownika (*user experience* – UX). Użytkownicy Facebooka stają się dostarczycielami treści, wymieniając informacje z życia prywatnego, dzieląc się różnymi materiałami⁶. Podstawową formą komunikacji są posty na „tablicy”, wiadomości wysyłane wewnętrznym komunikatorem lub komentarze pozostawiane pod zamieszczonym zdjęciem lub filmem wideo⁷. Poza określeniem *social media* inna często używana w tej publikacji nazwa to „platformy internetowe”, rozumiana jako „miejsce spotkań (...), wielka agora interakcji

³ A. M. Kaplan, M. Haenlein, *Users of the world, unite! The challenges and opportunities of Social Media*, „Business Horizons” 2010, no. 53, s. 59–68.

⁴ M. Juza, *Między wolnością a nadzorem. Internet w zmieniającym się społeczeństwie*, Scholar, Warszawa 2019, s. 223.

⁵ N. Dogruer, I. Meneviş, R. Eyyam, *What is the motivation for using Facebook?*, „Procedia – Social and Behavioral Sciences” 2011, no. 15, s. 2642–2646, za: M. Iwanowska, *Zrozumieć media (społecznościowe). Perspektywa psychologiczna*, w: *Media, biznes, kultura. Rzeczywistość medialna. Formy, problemy, aspiracje*, red. M. Łosiewicz, A. Ryłko-Kurpiewska, Novae Res, Gdynia 2016, s. 169–185.

⁶ M. Castells, *Władza komunikacji*, PWN, Warszawa 2013.

⁷ A. Gralczyk, *Współczesne portale społecznościowe jako nowa forma komunikacji*, w: *Media audiowizualne i cyfrowe wobec wyzwań współczesnego społeczeństwa*, red. A. Adamski, A. Gralczyk, K. Kwasik, M. Laskowska, Warszawa 2012, s. 203–204, za: K. Forst, *Facebook we współczesnym świecie komunikacji społecznej*, w: *Media społecznościowe: dialog w cyberprzestrzeni*, red. K. Doktorowicz, t. 2, Wydawnictwo Uniwersytetu Śląskiego, Katowice 2017, s. 37–55.

i transakcji, łączenia ludzi, firm i zasobów”⁸ i „programowalna architektura cyfrowa zaprojektowana do organizowania interakcji między użytkownikami, (...) ukierunkowana na systematyczne gromadzenie, przetwarzanie algorytmiczne, monetyzację danych użytkowników”⁹.

Wprowadzenie pojęcia sztucznej inteligencji do naukowej dyskusji przypisuje się Johnowi McCarthy’emu. W pierwotnym rozumieniu termin ten miał oznaczać „[proces, który] sprawia, że maszyna zachowuje się w sposób, który nazwalibyśmy inteligentnym, gdyby w ten sposób zachowywał się człowiek”¹⁰. Takie rozumienie dominowało w latach 50. XX wieku. Zaawansowanie prac nad systemami obliczeniowymi, które za pomocą danych podejmują decyzje lub pokazują sieci korelacji, rozwinęło się w Stanach Zjednoczonych i stamtąd przyszło do Europy i Azji. Innowacyjne podejście do technologii kopiującej sieci neuronowe ludzkiego mózgu nastąpiło wskutek rozwoju internetu i dostępu do dużej liczby danych (ang. *big data*), gdzie dane cyfrowe umieszczane przez użytkowników karmiły uczący się system. Dopiero odkrycie Geoffreya Hintona w pierwszej dekadzie XXI wieku pomogło w rewolucji technologicznej, która obrazowo przypominała „nafaszerowanie starych sieci sterydami”¹¹, w efekcie czego doszło do zwielokrotnienia mocy rozwiązywania problemów takich jak rozpoznawanie mowy czy twarzy. Nową zdolność komputera nazwano „uczeniem głębokim”, a kropkę nad i postawiła sieć neuronowa Hintona, która w roku 2012 pokonała rywali w międzynarodowych zawodach widzenia komputerowego.

Od kilku lat mamy już w Polsce do czynienia z początkiem debaty publicznej nad prawnym statusem sztucznej inteligencji, wiele firm wykorzystuje w praktyce bardziej proste lub zaawansowane jej formy w zarządzaniu i marketingu. Liczne przestrzenie życia społecznego czeka interwencja algorytmów, które będą wspierały medycynę, infrastrukturę, edukację, sektor usług. Unia Europejska określiła status sztucznej inteligencji jako:

⁸ J. Krefit, *Władza platform. Za fasadą Google, Facebooka i Spotify*, Universitas, Kraków 2022, s. 25.

⁹ V. Dijck, T. Poell, M. de Waal, *The platform society. Public values in a connective world*, Oxford 2018, za: J. Krefit, dz. cyt., s. 19–20.

¹⁰ M. Tegmark, *Życie 3.0. Człowiek w erze sztucznej inteligencji*, Prószyński i S-ka, Warszawa 2019, s. 60.

¹¹ K.-F. Lee, *Inteligencja sztuczna, rewolucja prawdziwa. Chiny, USA i przyszłość świata*, Media Rodzina, Poznań 2023, s. 22.

oprogramowania komputerowe stworzone przez człowieka, które, biorąc pod uwagę założony cel, działają w wymiarze fizycznym lub cyfrowym poprzez postrzeganie ich otoczenia dzięki gromadzeniu danych, interpretacji zebranych ustrukturyzowanych lub nieustrukturyzowanych danych, rozumowaniu na podstawie wiedzy lub przetwarzaniu informacji pochodzących z tych danych oraz podejmowaniu decyzji w sprawie najlepszych działań, które należy podjąć w celu osiągnięcia określonego celu¹².

Zapowiedzi platform internetowych o moderowaniu treści przy zastosowaniu sztucznej inteligencji nadal wzbudzają zaciekawienie obserwatorów życia społecznego, prowokują pytania o szanse i zagrożenia z nią związane.

Sztuczna inteligencja jest w tej publikacji rozumiana wąsko, jako program komputerowy, skoncentrowany na konkretnych czynnościach, mający na celu detekcję i eliminację danych. Zjawisko mowy nienawiści w internecie opisywane przez badaczy nauk społecznych zawierało początkowo takie synonimy jak: agresja elektroniczna¹³, agresywny wpis, nienawistny mem, bluzg¹⁴. Mówienie o nienawiści jako zwerbalizowanej emocji jest hiperbolą, bo definicje są wyolbrzymione semantycznie i nie chodzi w nich o „pragnienie niszczenia i wywrotową negatywność” – bardziej o „uczucie wrogości”¹⁵. Obecne w dyskursie naukowym pojęcia nie uwzględniają filozoficznych odniesień, które nienawiść definiują jako wzgardę dla słabych, chorych i biednych lub trujących od środka resentment¹⁶. Eksperti od retoryki i politologii analizujący agresję językową w polskiej debacie publicznej piszą o retoryce nienawiści, która przejawia się w ośmieszaniu, degradowaniu i dyskredytacji

¹² Definicja Sztucznej Inteligencji: Główne funkcje i dyscypliny, Grupa ekspertów wysokiego szczebla ds. sztucznej inteligencji, Bruksela: Komisja Europejska 2019, s. 6, https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60436.

¹³ J. Pyżalski, *Agresja elektroniczna i cyberbullying jako nowe ryzykowne zachowania młodzieży*, Impuls, Kraków 2012.

¹⁴ M. Juza, *Hejterstwo w komunikacji internetowej: charakterystyka zjawiska, przyczyny i sposoby przeciwdziałania*, „Profilaktyka Społeczna i Resocjalizacja” 2015, nr 25, s. 30.

¹⁵ B. Żmuda, *Czy może istnieć nienawiść usprawiedliwiona? Problem nienawiści ofiar do oprawców w myśli filozoficznej Barbary Skargi*, w: *Nienawiść w życiu publicznym*, red. W. Zuziak, J. Synowiec, Wydawnictwo Naukowe Uniwersytetu Papieskiego Jana Pawła II w Krakowie, Kraków 2016, s. 42.

¹⁶ M. Żardecka-Nowak, *Resentiment jako postać nienawiści*, w: *Nienawiść w życiu publicznym*, dz. cyt., s. 59.

przeciwnika¹⁷ oraz przypisywaniu osobom lub grupom szczególnie negatywnych cech lub uczynków¹⁸.

Pionierskie badania nad mową nienawiści w Polsce pokazują skrajne stanowiska. Językoznawcy określają nią wulgarne zwroty służące obrażaniu konkretnych grup, a do jej przyczyn zaliczają chęć deprecjacji, frustrację, potrzebę zauważenia¹⁹ oraz patologicznie rozumianą wolność słowa²⁰. W opinii psychologów źródłem mowy nienawiści jest pogarda i zepsucie życia politycznego w Polsce²¹ oraz kulturowy zanik prywatności i medialny ekshibicjonizm, będący efektem ubocznym cyfryzacji²². Zdaniem badaczy hejterstwo jest inspirowane przez właścicieli platform, ponieważ nienawistne komentarze napędzają dyskusję i zwiększają zyski²³. Diagnozuje się społeczne przyzwolenie na hejt w internecie i pojawienie się wielu jego odmian, np. „hejtu swojskiego”²⁴, będącego konsekwencją wojny polsko-polskiej, prowadzonej przez polityków nie tylko podczas kampanii wyborczych. Od kilku lat dyskutowany jest temat przeciwdziałania mowie nienawiści, tak w środowisku akademickim, jak technologicznym, i poglądy się ścierają. Jedna strona mówi o moderacjach w cyberprzestrzeni i blokadzie obraźliwych treści, druga – o silnej penalizacji hejterów. Powtarzającym się postulatem jest edukacja społeczeństwa nt. odpowiedzialności za słowa w przestrzeni on-line, gdzie anonimowość i bezkarność mają złudny charakter.

¹⁷ M. Głowiński, *Retoryka nienawiści*, „Nauka” 2007, nr 2, s. 24.

¹⁸ L. Nijakowski, *Mowa nienawiści w świetle teorii dyskursu*, w: A. Horolets, *Analiza dyskursu w socjologii i dla socjologii*, Adam Marszałek, Toruń 2008, s. 113.

¹⁹ A. Naruszewicz-Duchlińska, *Nienawiść w czasach internetu*, Novae Res, Olsztyn 2015, s. 35.

²⁰ M. Osowiecka, *Nienawistny język w Internecie*, „Ja. My. Oni. Poradnik Psychologiczny Polityki”, 26.04.2015, t. 18, <https://www.polityka.pl/jamyoni/1617093,1,nienawistny-jezyk-w-internecie.read> (11.10.2023).

²¹ *Więcej niż polityka*, Michał Bilewicz w rozmowie z Danielem Lisem, „Znak” 2019, nr 4, s. 27.

²² A. Frącz, *Hejting – przejawy mowy nienawiści w Sieci*, „Parezja” 2019, nr 2, s. 30.

²³ M. Juza, *Hejterstwo...*, art. cyt., s. 27.

²⁴ A. Leszczuk-Fiedziukiewicz, *Inicjatywy antyhejtowe. Wybrane przykłady zaangażowania społecznego w ograniczanie mowy nienawiści*, w: *Język a media. Perspektywy i zagrożenia języka we współczesnych mediach*, red. E. Horyń, B. Skowronek, A. Walecka-Rynduch, Collegium Columbianum, Kraków 2020, s. 102.

Czy sztuczna inteligencja będzie moralnym strażnikiem społeczeństwa?

Josh Simons i Eli Frankel, młodzi badacze technologii i jej wpływu na demokrację, zapowiadając: „cały sektor usług woła o dobrą predykcję i dyskryminację, czyli stosowanie metod EDA [eksploracyjna analiza danych]”²⁵, zauważyli wyzwanie, przed jakim stoją władze państwowe i organizacje biznesowe, które już niebawem będą wykorzystywać sztuczną inteligencję. Naukowcy prognozują czasy, w których sztuczna inteligencja będzie zajmować się moderowaniem sfery publicznej. Ich zdaniem potrzebna będzie refleksja nad funkcjami, jakie sztuczna inteligencja ma pełnić w zależności od przestrzeni życia społecznego – inne w polityce, inne w mediach społecznościowych. Według Simonsa i Frankela kontakt ze sztuczną inteligencją wymusi na środowisku akademickim, politycznym i biznesowym trudne pytania o kształt sfery publicznej. Dotyczą one roli człowieka w zarządzaniu technologią tak, by nie była ani powodem manipulacji i powielania kłamstw, ani narzędziem destrukcyjnej walki politycznej.

Raporty organizacji pełniących funkcję „watchdogów”, patrzących na ręce właścicieli mediów społecznościowych, donoszą o efektach działań sztucznej inteligencji stosowanej przez Facebook i Twitter. Zamknięcie w ciągu dwóch miesięcy 1,5 mld kont na Facebooku i 70 mln kont na Twitterze ze względu na dezinformację i mowę nienawiści²⁶ to wynik zaskakujący, jednak warto pamiętać o globalnym wymiarze tych marek. Analizujący ów raport dziennikarze „The Wall Street Journal” zadają pytanie o to, czy algorytmy staną się niebawem moralnym strażnikiem komunikacji między ludźmi²⁷. To pytanie nie jest bez sensu, ponieważ technologiczne boty będą wkrótce zarządzać wieloma sferami życia społecznego, także emocjami ukrytymi w słowach pisanych on-line. Zapowiedzi platform internetowych

²⁵ J. Simons, E. Frankel, *Dlaczego sztuczna inteligencja powinna być demokratyczna?*, Wszystko co Najważniejsze, 17.03.2023, <https://wszystkoconajwazniejsze.pl/josh-simons-i-eli-frankel-sztuczna-inteligencja-demokratyczna/> (10.08.2023).

²⁶ *Twitter usunął 70 mln fałszywych kont*, 10.07.2018, <https://www.euractiv.pl/section/gospodarka/news/twitter-usunal-70-mln-falszywych-kont-w-dwa-miesiace/> (10.08.2023).

²⁷ W. Orzeł, *Czy sztuczna inteligencja poradzi sobie z mową nienawiści?*, <https://aibusiness.pl/czy-sztuczna-inteligencja-poradzi-sobie-z-mowa-nienawisci/> (10.10.2023).

o moderowaniu treści przy wykorzystaniu sztucznej inteligencji wzbudzają zaciekawienie obserwatorów życia społecznego, prowokują pytania o szanse i zagrożenia z nią związane.

Ewolucja polityki Facebooka wobec mowy nienawiści

Żyjemy w czasach triumfalizmu Facebooka, który jak Feniks z popiołów, mimo kompromitujących posunięć z przeszłości, nadal odgrywa ważną rolę w życiu społecznym. Widzimy, jak młode pokolenia dokonują sakralizacji tego medium, obserwujemy gigantomanię i fetyszyzację Facebooka w globalnej gospodarce. Hegemonię na rynku nowych mediów (ponad 3 mld użytkowników w roku 2023) potwierdziło również grono polityków, które zapraszało najpierw Marka Zuckerberga, potem Elona Muska na polityczne salony. Rozmowy o negatywnych treściach mających wpływ na nastroje różnych grup społecznych, i wymagających aktywnej moderacji, przeprowadził z szefami Facebooka i Barack Obama, i Emmanuel Macron. Konflikty światowe skurczyły się do rozmiarów postów i tweetów. Kilka zdań może wywoływać rewolucje.

Jednym z istotnych pytań jest to, w jaki sposób ludzie zatrudnieni w Facebooku podchodzili do niewygodnych treści pojawiających się na stronach portalu, zanim zaczęto w nim stosować sztuczną inteligencję. Najpierw działania te miały charakter amatorski, potem profesjonalny. Za pomocą metod hakerskich inżynierowie młodego start-upu z Doliny Krzemowej, włamując się na konto osoby zgłaszającej, sprawdzali jej wiarygodność, potem tak samo wchodzili na konto osoby zgłaszanej i ręcznie usuwali zgłoszony wpis. Warto przyjrzeć się zjawisku moderowania treści przez administrację największego komunikatora, odwiedzanego codziennie przez miliardy ludzi na całym świecie. W 2009 r. w Dublinie Facebook utworzył biuro odpowiedzialne za szkolenie moderatorów i usuwanie negatywnych wpisów pojawiających się na stronach. Corocznie, pod naciskiem opinii publicznej, zwiększał grupę pracowników moderujących niewygodne treści. W 2019 r. na zlecenie Facebooka internet czyściło już 15 tysięcy moderatorów, i byli

to głównie pracownicy zatrudnieni przez firmy outsourcingowe na Filipinach i w Indiach. Z relacji dziennikarzy wiemy, że ich praca trwała 24 godziny na dobę, a każdy z moderatorów miał ok. 40 sekund na decyzję, czy dany post lub nagranie należy usunąć. Praca polegała na oglądaniu i selekcji okrutnych scen, np. wykorzystywania seksualnego dzieci, odcinania głów i transmisji ataków terrorystycznych. Levy pisał o nich:

[...] stali się odpowiednikami cyfrowych dozorców, czyszczących News Feeda niczym niewidzialni sprzątacze, którym płaci się za oglądanie takich treści jak sceny gwałtów, nielegalne operacje medyczne, ludzkie genitalia (...), a praca rzuca im się na mózg²⁸.

Inni informatycy mówili o nich: „szybko się wypalali, (...) zmuszeni byli patrzeć na to, co najgorsze w człowieku”²⁹. Wskutek kontaktów ze wstrząsającymi treściami Facebook zaczął zapewniać moderatorom pomoc psychoterapeutów. Wielu z nich miało objawy bliskie stresu pourazowego, znanego jako PTSD, lub inne zaburzenia psychiczne³⁰, opisane przez portale „The Verge” i „The Financial Times”. Facebook został zmuszony przez sąd do zapłaty odszkodowań 11 tysiącom moderatorów za toksyczne warunki pracy, mającej destrukcyjny wpływ na psychikę. Kuriozalny wydaje się fakt, że z zasądzonej w maju 2020 r. kwoty 52 milionów dolarów poszkodowani otrzymali jedynie 28 tysięcy dolarów³¹.

Nagłośnienie kampanii dezinformacyjnej przygotowanej przez fabryki rosyjskich trolli z Sankt Petersburga doprowadziło do głośnego procesu prowadzonego przez Kongres USA w 2016 r. Po wyroku i rekordowej karze

²⁸ S. Levy, *Facebook. A miało być tak pięknie. Za kulisami imperium z Doliny Krzemowej*, tłum. K. Sosnowska, Białystok 2022, s. 465-466.

²⁹ W. Orzeł, *Czy sztuczna inteligencja poradzi sobie z mową nienawiści*, dz. cyt.

³⁰ *Traumatized Facebook moderators turn to sex on the job*, „New York Post”, 25.02.2019, <https://nypost.com/2019/02/25/trauma-a-job-hazard-for-facebook-content-moderators/> (12.10.2023); *Facebook and YouTube moderators sign PTSD disclosure*, BBC, 25.01.2020, <https://www.bbc.com/news/technology-51245616> (12.10.2023); *Facebook zapłaci 52 mln dolarów odszkodowania moderatorom za pracę w trudnych warunkach*, <https://www.wirtualnemedial.pl/arttykul/facebook-zaplaci-52-mln-dolarow-odszkodowania-moderatorom-za-prace-w-trudnych-warunkach> (13.10.2023).

³¹ J. Kreft, *Władza platform*, dz. cyt.

5 mld dolarów³² Facebook pod presją opinii publicznej musiał pokazać światu (reklamodawcom i użytkownikom), że jest firmą odpowiedzialną społecznie. W procesie oskarżającym Facebook o niereagowanie na dezinformację ze strony Rosji Mark Zuckerberg mówił: „jeśli chodzi o mowę nienawiści jestem optymistą, że w ciągu 10 lat narzędzia AI będą rozróżniać detale językowe, ale jest to teraz niemożliwe i muszą zajmować się tym ludzie”³³. Kongresmenom i opinii publicznej obiecywał też zwiększyć zespół moderacji treści do 20 tysięcy ludzi, którzy będą ręcznie usuwali obraźliwe wypowiedzi i obsceniczne posty. Po wizerunkowej kompromitacji Zuckerberg z zespołem stworzyli *Standardy społeczności Facebooka* i konsultowali je z przedstawicielami różnych środowisk, krajów i kultur, którzy spotykali się w szczytnym celu stworzenia bazy słów i tematów objętych rygiem. Ich namierzanie i kasowanie Facebook nazwał działaniem proaktywnym. Praktyka pokazywała wiele paradoksów tych reguł, gdyż Zuckerberg nie reagował na wpisy Donalda Trumpa i tworzenie podziałów wśród użytkowników³⁴. Zarzucano mu ignorowanie agresji słownej prowadzącej do ludobójstwa. W raporcie firm BSR i ONZ pisano „Facebook stał się użyteczną platformą dla tych, którzy chcą podżęgać do przemocy i krzywdzić poza siecią”³⁵. Koordynatorka działań firmy w zakresie wolności słowa, Monika Bickert, przyznała: „polityka w sprawie nienawiści jest naszym najtwardszym orzechem do zgryzienia, ponieważ brakuje nam kontekstu”³⁶. Zdaniem znawców³⁷ sama struktura platformy pozwala na praktyki plemienne, sianie niezgody i organizację kampanii nienawiści.

³² J. Krawczyński, *Rząd USA nałożył 5 mld kary dla Facebooka. Firma na tym zyskała*, 14.07.2019, <https://www.purepc.pl/rzad-usa-nalozyl-5-mld-kary-dla-facebook-a-firma-na-tym-zyskala> (10.12.2023).

³³ K. Winczewska-Latoszek, *Wyzwania i zagrożenia sztucznej inteligencji*, <https://web.swps.pl/strefa-zarzadzania/artykuly/19084-to-ja-robot-wyzwania-i-zagrozenia-sztucznej-inteligencji?dt=1704380584740> (13.10.2023).

³⁴ Sh. Frenkel, C. Kang, *Brzydka prawda. Kulisy walki Facebooka o dominację*, Rebis, Poznań 2022, s. 24.

³⁵ S. Levy, dz. cyt., s. 458.

³⁶ Tamże, s. 469.

³⁷ R. McNamee, *Nabici w Facebooka. Przestroga przed katastrofą*, tłum. K. Puławski, Warszawa 2019, s. 246.

Uczenie maszynowe narzędziem komunikacyjnym Facebooka

Pretekstem do kolejnego kroku w zarządzaniu mową nienawiści była sprawa skandalu z Cambridge Analytica. Określono tak wyciek danych użytkowników bez ich zgody i udostępnienie ich firmie przygotowującej profilowane reklamy polityczne, które pomogły Donaldowi Trumpowi wygrać wybory prezydenckie w USA. Sprawa ta przyniosła dużą nieufność wobec Facebooka i rezygnację wielu użytkowników z jego usług. Na ratunek zrujnowanej opinii ruszył zespół Responsible AI, kierowany przez Joaquina Candele, w branży IT określony jako zautomatyzowana policja myśli³⁸, a także nowe działy ds. uczciwości i bezpieczeństwa. Innowacja Facebooka w zastosowaniu „odpowiedzialnej sztucznej inteligencji” (ang. *responsible artificial intelligence*) polegała na zastosowaniu tych samych algorytmów, zarówno do pobudzania ruchu na stronach, jak i do ograniczania niepożądanych treści. Doszło do tego stworzenie pracowni Sztucznej Inteligencji pod kierownictwem francuskiego naukowca Yanna LeCuna, który miał do swojej dyspozycji zespół AI Research³⁹. Przy angażowaniu sztucznej inteligencji okazywało się, że barierą nie do przeskoczenia były aspekty kulturowe, typowe dla określonych części świata, oraz kontekst komunikacji, czyli słowa zabarwione emocjonalnie w danym języku lub dialekcie. Maszyna miała odróżniać brutalne starcia i nieszkodliwe sprzeczki, jednak w praktyce nie radziła sobie z detekcją agresji. Udowodniły to pierwsze próby algorytmu, gdy wpisy dotyczące samochodowych wypadków sztuczna inteligencja kwalifikowała jako walki kogutów⁴⁰. Przypadki te były nagłaśniane dopiero po pewnym czasie lub wyciekały dzięki medialnym skandalom. Jednym z nich było wystąpienie Francis Hagen, twierdzącej, że Facebook używa

³⁸ *Internetowi potentaci razem przeciw mowie nienawiści, Facebookowi pomoże w tym sztuczna inteligencja*, <https://www.dobreprogramy.pl/internetowi-potentaci-razem-przeciw-mowie-nienawisci-facebookowi-pomoze-w-tym-sztuczna-inteligencja,6628359852615809a> (30.09.2023).

³⁹ K.-F. Lee, *Inteligencja sztuczna, rewolucja prawdziwa. Chiny, USA i przyszłość świata*, tłum. K. Hejwowski, Media Rodzina, Poznań 2019, s. 115–117.

⁴⁰ M. Fraser, *Moderacja treści na Facebooku oparta o sztuczną inteligencję nie działa*, 18.10.2021, <https://cyberdefence24.pl/social-media/moderacja-tresci-na-facebooku-oparta-o-sztuczna-inteligencje-nie-dziala> (30.09.2023).

algorytmu sprzyjającego kłótniom i podziałom użytkowników. Jej tezy koncentrowały się na algorytmie, który został wymyślony tak, by prowokować na stronie ruch tylko tych informacji, które zawierają w sobie treści prowokacyjne. System platformy izolował w konsekwencji użytkowników (członków plemion), zamykając ich w bańkach informacyjnych (filtrujących i preferencyjnych⁴¹), a bycie częścią plemienia zamykało na inne wiadomości w myśl zasady: dzielimy się z przyjaciółmi naszymi treściami i czytamy ich treści, bo im ufamy, nawiązujemy relację tylko z tymi, którzy podzielają naszą wizję świata. Hagen mówiła: „Facebook świadomie szkodzi, algorytmy celowo zaogniają podziały społeczne w USA, nie są apolityczne”⁴². W jej opinii dochodziło do tendencyjnego kształtowania opinii publicznej, a platforma stała się „miejszem bardziej gniewnym”. Lars Backstrom, wiceprezes ds. inżynierii, odpowiedział na zarzuty sygnalistki, tłumacząc się, że każdy algorytm ryzykuje promowanie treści, które są szkodliwe dla niektórych użytkowników, Chris Cox z kolei tłumaczył: „im więcej zaangażowania, tym lepiej one działają!”⁴³. Po nagłośnieniu afery duża liczba użytkowników i reklamodawców odeszła z Facebooka, który zaczął być kojarzony z nieuczciwością i inżynierią społeczną.

Wyzwanie stojące przed Facebookiem to wyzwanie karkołomne; przy jednoczesnym zachowaniu statusu neutralnego kanału komunikacji (Zuckerberg od początku nie chce nazywać Facebooka medium, tylko platformą udostępniającą ludziom miejsce do porozumiewania się) istnieje duże oczekiwanie społeczne, by władze portalu reagowały na wylew nienawistnych komentarzy i nieetyczny język prowokujący do przemocy, zwłaszcza że zdaniem badaczy informacje na stronach Facebooka mogą służyć prześladowaniu i upokarzaniu⁴⁴. Pytanie, czy Facebookowi uda się określić jeden

⁴¹ Homofilia jest konsekwencją działań platform i wyborów internautów, zjawisko to polega na wyodrębnianiu się wspólnot (plemion według Maffesollego) złożonych z osób tak samo myślących i zamkniętych na inne punkty widzenia, M. Juza, *Między wolnością a nadzorem*, art. cyt., s. 228.

⁴² *Amerykańskie media alarmują: Facebook przyczynił się do „społecznej wojny domowej” w Polsce*, Polskie Radio, 30.10.2021, <https://www.polskieradio.pl/129/1280/Artykul/2837300%2CAmerykanskie-media-alarmuja-Facebook-przyczynil-sie-do-spolecznej-wojny-domowej-w-Polsce> (30.08.2022).

⁴³ S. Levy, dz. cyt., s. 454.

⁴⁴ S. Vaidhyanathan, *Antisocial Media. Jak Facebook oddala nas od siebie i zagraża demokracji*, Warszawa 2018, s. 90-91.

uniwersalny język nienawiści, którego nauczy swoje algorytmy, jest pytaniem ważnym i aktualnym.

Przyglądanie się temu, jak Facebook radzi sobie z obietnicami reagowania na obsceniczne i wulgarne treści, jest zajęciem wielu organizacji monitorujących media. Podczas gdy firma przekonywała, że 97% usuniętych treści jest identyfikowane przez algorytm, obserwatorzy zarzucali jej nieskuteczność stosowanej w tym procesie sztucznej inteligencji. Wiceprezes Facebooka podawał do opinii publicznej liczbę wykrywalności hejtu on-line w wysokości 50%, dziennikarze śledczy z redakcji „Wall Street Journal”, bazując na wypowiedziach pracowników Facebooka, opublikowali raport o niskim, jednocyfrowym odsetku wpisów naruszających zasady postępowania. Gazeta podawała też koszty zatrudnienia moderatorów: 2 mln dolarów tygodniowo i 104 mln rocznie, którymi firma uzasadniała zatrudnienie maszynowych algorytmów do działań, którym nie jest w stanie podołać grupa tysięcy ludzi. Kiedy indziej Facebook ogłaszał swoje sukcesy w stosowaniu sztucznej inteligencji, chwając się, że 88,8% mowy nienawiści jest wykrywane prewencyjnie⁴⁵, podczas gdy jego oponenti twierdzili, że Facebook ma „cenzorskie zapędy” (blokowanie lub blokowanie dostępu do stron⁴⁶), zwłaszcza w przypadku treści konserwatywnych lub materiałów nieaprobowanych przez WHO podczas pandemii. Jednak czy zablokowanie dwóch miliardów prób otwarcia nowych kont w 2019 r. i skasowanie 9,6 mln wpisów przez pierwsze trzy miesiące roku 2020 można odebrać jako sukces? Warto pamiętać, że użycie algorytmów do zarządzania niepożądanymi, szkodliwymi treściami jest uzasadniane dezinformacją, będącą konsekwencją możliwości technicznych social mediów. Zakładanie fałszywych kont i rozsyłanie nieprawdziwych informacji po rozpoczęciu przez Rosję wojny na Ukrainie i prowadzeniu antyukraińskiej propagandy ma coraz większe rozmiary. Jak widać, Facebook ma swoją własną historię w walce z mową nienawiści, która pojawia się na jego stronach. Jego doświadczenia nie mogą być przykładem do naśladowania, ponieważ zarządzający platformą, goniąc za innowacyjnością, podjęli wiele błędnych decyzji, które zmniejszyły stopień zaufania społecznego wobec ich działań.

⁴⁵ *Sztuczna inteligencja będzie cenzorem na FB*, „Frona”, 19.05.2020; <https://www.frona.pl/a/dzieki-nowym-rozwiazaniom-si-cenzura-na-facebooku-bedzie-skuteczniejsza,144531.html>.

⁴⁶ Tamże.

Rasistowskie przypadki Twittera

Twitter to kolejna platforma internetowa służąca masowej komunikacji, dająca możliwość umieszczania krótkiej wiadomości tekstowej w przestrzeni on-line. Podobnie jak Facebook, zadeklarowała, że walczy z mową nienawiści za pomocą sztucznej inteligencji. Firma chciała udowodnić, że jest organizacją odpowiedzialną społecznie, ponieważ była bohaterem skandalu z powodu treści antysemitycznych. Posty wychwalające Adolfa Hitlera pojawiały się obok reklam Apple, Oracle czy IBM, co nie spodobało się firmom kupującym miejsca reklamowe i spowodowało w środowisku biznesowym stygmatyzację Twittera⁴⁷. Zarzucano mu, że wykorzystywane algorytmy nie blokują rasistowskich treści, i zaczęto rezygnować z reklam. Takie wpadki z mową nienawiści były prawdopodobnie przygotowaną specjalnie akcją propagandową dyskredytującą Twittera; mogły to zrobić firmy konkurencyjne, wykorzystujące hacking w ramach tzw. czarnego PR-u. O niszczeniu reputacji firmy przez konkurencję jako często spotykanej patologii mówi się w środowisku biznesowym dość głośno. Opisane wpadki mogły być również konsekwencją trollingu, znanego też w innych social mediach.

Temat sztucznej inteligencji jest zajmujący także dlatego, że algorytmy zarządzające treściami w internecie weszły na salony polityczne. W 2022 r. Elona Muska, nowego właściciela Twittera, zaprosił na rozmowę sam prezydent Francji, Emmanuel Macron. Spotkanie dotyczyło modyfikacji treści za pomocą sztucznej inteligencji. Musk deklarował potem opinii publicznej skuteczność działań, mówiąc, że „liczba interakcji wobec tweetów nawołujących do nienawiści spada mimo znaczącego wzrostu liczby użytkowników”. Na kłamstwie przyłapał Muska dziennikarz „The New York Timesa”, który twierdził, że odkąd Musk przejął platformę, skala mowy nienawiści wzrosła 500-krotnie. Podawał też liczby: wzrost liczby postów antysemitycznych o 60%, częstsze wpisy o homoseksualistach: „kiedyś było ich 2506 dziennie, teraz [grudzień 2022] jest prawie 4 tysięcy

⁴⁷ *Disney, IBM, Apple i inne marki wycofują reklamy z X/Twittera. Powód: antysemityczne treści*, <https://www.wirtualnemedi.pl/arttykul/koniec-reklam-x-twitter-disney-ibm-apple-antysemityzm> (30.08.2023).

dziennie⁴⁸. Po zarzutach nowojorskiej redakcji głos zabrała szefowa Twitera ds. zaufania i bezpieczeństwa, Ella Irwin, która twierdziła, że Twitter usunął w tamtym czasie 44 tysiące kont związanych z naruszeniem bezpieczeństwa dzieci. Zauważalna jest podobna strategia komunikacji co u Facebooka, gdy platforma, będąc oskarżana o udostępnianie miejsca nienawistnym wpisom lub niereagowanie na nie, pokazuje inne etyczne działania na rzecz użytkowników, chociażby związane z ochroną dzieci.

Co ciekawe, niektórzy naukowcy obserwujący rozwój mediów społecznościowych wytykają niekonsekwencję w decyzjach ich właścicieli. Elon Musk jeszcze kilka lat temu przestrzegał przed ekspansją sztucznej inteligencji, dziś chwali się współpracą z izraelskim start-upem, który ma reagować na mowę nienawiści. Twitter nagłaśniał zakup dwóch start-upów, Madbits i Whetlab⁴⁹, które zdaniem Muska zaprogramowały sztuczną inteligencję myślącą się tylko w 7% przypadków. Deklaracje najbogatszych ludzi świata zwykle mają wymiar marketingowy. Wzrost liczby osób tweetujących oznacza wzrost problematycznych treści. Musk jednak nie traci rezonu, podkreśla innowacyjność wymyślonego przez start-up programu, chwając się, że sztuczna inteligencja umie odczytać sarkazm i sama ma poczucie humoru. Patrzący mu na ręce obrońcy bezpiecznego internetu i dziennikarze śledczy twierdzą, że Elon Musk mija się z prawdą, bo ich zdaniem sytuacja wymyka się spod kontroli, a Twitter po masowych zwolnieniach moderatorów jest pełen treści antysemitycznych. Z kolei inni obserwatorzy twierdzą, że systemy sztucznej inteligencji pozwoliły ograniczyć liczbę ludzkich moderatorów, ale nie rozwiązały kwestii ręcznego zawieszania niewygodnych kont.

⁴⁸ Tamże.

⁴⁹ *Mowa nienawiści na Twitterze. Sztuczna inteligencja ma wspierać moderację*, <https://cyberdefence24.pl/social-media/mowa-nienawisci-na-twitterze-sztuczna-inteligencja-ma-wspierac-moderacje> (30.08.2023).

Rozdźwięk między presją społeczną a brakiem zaufania do „gigantów”

Zaangażowania w kasowanie hejtu z mediów społecznościowych oczekują politycy i zwykli ludzie. Facebook i Twitter odpowiadają na te oczekiwania w sposób marketingowy, deklarując działania i zaangażowanie, odwołując się do szeroko rozumianego bezpieczeństwa użytkowników. Zanim deklarowano reakcję na hejt w internecie, udzielono przyzwolenia na kontrolowanie zachowań internautów ze względu na postulaty walki z „czterema jeźdźcami infokalipsy”: praniem brudnych pieniędzy, narkotykami, terroryzmem i pornografią dziecięcą⁵⁰. W 2016 r., wraz z innymi firmami technologicznymi, Facebook podpisał zobowiązanie do walki z mową nienawiści, polegające na usuwaniu nienawistnych treści w ciągu 24 godzin od otrzymania zgłoszenia. Zdaniem komentatorów porozumienie, jakie podpisali z Komisją Europejską amerykańscy giganci technologiczni, nie wynikało z pobudek etycznych, lecz z obawy przed zastosowaniem kar finansowych wobec tych, którzy zakazanych treści by nie usunęli. Porozumienie nazwane *Kodem postępowania*⁵¹ zaistniało w opinii publicznej jako ważny krok w normalizacji przestrzeni on-line i przeciwdziałanie szkodliwym treściom. Z drugiej strony raporty unijne potwierdzają spadek liczby reakcji platform, bo w 2020 r. rozpatrywały one 90% zgłoszeń⁵², w 2022 już 64%, podczas gdy chiński TikTok poprawił wyniki, usuwając 90% zgłoszonych treści. Widać, że w wielu wymiarach konkurencja technologii trwa.

Raport unijnych urzędników pokazywał, że najczęściej usuwane były treści wzywające do zabójstwa, rzadziej treści zniesławiające kogoś. Za bardziej wiarygodnych od użytkowników indywidualnych zgłaszających hejt firmy uznawały zaufane organizacje monitorujące internet. Věra Jourová

⁵⁰ Mowa o pomysłach opisanym w książce *Cyberpunks. Wolność i przyszłość internetu*, autorstwa J. Assange’a, J. Appelbauma, A. Muller-Maguhna i J. Zimmermana, tłum. M. Michnik, Wyd. Helion, Gliwice 2013; za: M. Juza, *Między wolnością a nadzorem*, art. cyt., s. 250.

⁵¹ *Komisja Europejska zastrzega zasady moderacji treści*, 27.04.2023, <https://www.virtualnedia.pl/artukul/komisja-europejska-zaostrza-zasady-moderacja-tresci-facebook-tiktok-twitter-google-wikipedia-apple> (11.10.2023).

⁵² *Internet. Platformy odpuszczają usuwanie mowy nienawiści*, 26.11.2022, <https://www.dw.com/pl/internet-platformy-odpuszczaj%C4%85-usuwanie-mowy-nienawi%C5%9Bci/a-63899752>.

w komentarzach do tych danych mówiła: „mowa nienawiści w sieci jest plagą ery cyfrowej”⁵³, wypominając platformom niesłowność w obietnicach. Zapowiedziała, że Unia będzie się przyglądać usuwaniu „nielegalnych treści” przez Twitter, Google i Facebook, grożąc zarazem karami do 10% rocznych obrotów firm. A już w roku 2022 Komisja Europejska ostrzegła Facebook, że sama wprowadzi regulacje ograniczające mowę nienawiści na platformie. Zuckerberg zabrał wówczas głos, oświadczając, że „to przeszkodzi innowacjom i wolności słowa”⁵⁴, ale jego ludzie zajęli się tematem z obawy o kary i utratę już nadwątlonego prestiżu.

Technologiczni giganci, Twitter i Facebook, mają więc podobny problem. Czekają ich jeszcze większa liczba emocjonalnych wpisów na ich stronach, spowodowana zmianami społecznymi w wymiarze lokalnym i geopolitycznym. Nastroje społeczne ulegają polaryzacji. Internet sprzyja komunikacji wiralowej, więc skasowane wpisy lub nagrania wideo będą się pojawiały w kolejnych miejscach w sieci. Dochodzi do tego konkurencja ekonomiczna innych gigantów, spoza Ameryki i Europy. Chiński TikTok i inne komunikatory walczą o uwagę publiczności i mogą być zainteresowane rysą na reputacji firm im zagrażających. Wszyscy właściciele chwalą się używaniem algorytmów przesiewających treści uznane powszechnie za obraźliwe, jednak nie mówią o kularach swoich działań. Czy społeczność internautów może liczyć na odpowiedzialność? W tym roku reprezentanci środowisk patrzących na ręce właścicieli social mediów raportują, że moderatorzy treści nie nadążają kasować postów naruszających zasady społeczności wyznaczone przez administratorów tych platform; rzecz dotyczy ośmiu tysięcy wpisów dziennie. Ich decyzje bywają poza tym niekonsekwentne i alogiczne, czego doświadczyli również użytkownicy Facebooka w Polsce.

⁵³ Tamże.

⁵⁴ *Internet. Platformy odpuszczają usuwanie mowy nienawiści*, dz. cyt.

Paradoksy walki z mową nienawiści przy zastosowaniu sztucznej inteligencji

Trudno jest oszacować skalę problemu mowy nienawiści obecnej w internecie, wyrażanej z różnych pobudek i na różne potrzeby. Zdiagnozowanie, co należy uznać za słowa, które obrażają, upokarzają lub nawołują do przemocy (i czy są nimi wulgaryzmy, czasem wykropkowane, a czasem zastąpione gwiazdkami), jest kolejnym wyzwaniem. Znalezienie odpowiedniego rozwiązania, satysfakcjonującego użytkowników mediów i ich właścicieli, a także opinię publiczną reprezentującą wiele światopoglądów, mającą różnorodną wrażliwość, jest prawie niemożliwe. Dość utopijne jest myślenie o uporządkowaniu internetowych treści budzących spory i konflikty. Dzieje się tak z czterech powodów: po pierwsze, braku zaufania do technologicznych gigantów, właścicieli mediów społecznościowych; po drugie, z wykorzystywania natury internetu do produkowania i utrzymywania konfliktów, po trzecie, z istoty samej sztucznej inteligencji, po czwarte, z potrzeby etycznego podejścia właścicieli mediów i platform internetowych oraz ich prawników. Powody te będą omówione poniżej.

Przestaliśmy ufać takim gigantom jak Facebook w kwestii chronienia nas przed niebezpieczną treścią, ponieważ jego właściciele swoimi decyzjami nieraz już udowadniali nieetyczność działań. Właściciele mediów społecznościowych deklarują dbanie o bezpieczeństwo, ale w praktyce mają cele inne niż etyczne i nie ujawniają swoich porażek. Za nieuznawaniem Facebooka za autorytet w dziedzinie etyki i bezpieczeństwa przemawiają dwa argumenty: niereagowanie na mowę nienawiści obecną na ich stronach i sprzyjanie reżimom (w Indiach, w dawnej Birmie, gdzie wskutek elektronicznej nagonki doszło do bójek i zabójstw), o czym piszą McNamee i Levy. Sprawą bez precedensu było też pozwanie Facebooka przez dwóch naukowców z Etiopii i kenijską grupę praw człowieka (Katiba Institute). Zarzucili oni platformie niereagowanie jej moderatorów na brutalne treści z 2021 r., które doprowadziły do wojny domowej w Nairobi⁵⁵, i postulowali wypłacenie 2 miliardów dolarów na fundusz ofiar przemocy. Zarzut dotyczył

⁵⁵ *Facebook pozwalał na brutalne, nienawistne wpisy? Naukowcy i aktywiści chcą 2 mld dol. dla ofiar przemocy*, 15.12.2022, <https://www.wirtualnemedia.pl/arttykul/facebook-pozwany-mowa-nienawisci>.

niestarannego przygotowania algorytmów w celu identyfikacji postów zawierających nawoływanie do przemocy i odmowy usunięcia wpisów mimo ich zgłoszenia. To chyba pierwszy raz, gdy algorytm stał się bohaterem procesu sądowego. Facebook poza tym od lat jest oskarżany o dystansowanie się wobec odpowiedzialności za swoje decyzje, co pokazały przypadki z wyborami, Cambridge Analytica i algorytmem napędzającym nienawiść na stronach⁵⁶. Firma jest nadal krytykowana za słabe zainteresowanie konsekwencjami, jakie wywołują algorytmy wśród młodych użytkowników. Przedstawiciele aż 33 stanów USA w listopadzie 2023 r. złożyli pozew sądowy przeciwko Meta (będącej właścicielem Facebooka), oskarżanej o „hakowanie umysłów”, tworzenie uzależnień behawioralnych i wzrost problemów psychicznych u młodych ludzi. Prawdopodobnie decyduje o tym biznesowe podejście i „etyczna dezorientacja kultury technologicznej”⁵⁷, która wyrosła na fundamentach aktywności hakerów i innowatorów, marzących o byciu ponad państwem i prawem, mających ambicje ścigania się z konkurencją o uwagę użytkowników.

Firmy technologiczne odkryły swoją nieetyczną twarz, więc sprawy biorą we własne ręce użytkownicy angażujący sztuczną inteligencję w usuwanie mowy nienawiści. Przykładem są polscy badacze, marzący o nauczaniu programu empatii. Technologiczny start-up Samurai Labs wraz z badaczami z Uniwersytetu Warszawskiego (Centrum Badań nad Uprzedzeniami) wymyślił bota, który na hejt w sieci reaguje grzecznością i empatią. W artykule naukowym prof. Michał Bilewicz z zespołem opisał pierwsze efekty obniżania poziomu nienawistnych postów⁵⁸. Inny projekt, koordynowany przez dr. Jana Koconia z Politechniki Wrocławskiej, opiera się na założeniu, że algorytm można nauczyć wrażliwości i subiektywnej oceny, czyli przewidywania, jakie treści nie będą podobały się użytkownikom. Ambicją naukowca jest nauczanie algorytmu przewidywania treści, w oparciu o analizę podobieństwa

⁵⁶ B. Goczał, *Algorytm Facebooka nagradza nienawiść. Zuckerberg nie chce zmian, boi się odwrotu użytkowników*, 19.09.2021, <https://www.wirtualnemedial.pl/artykul/algorytm-facebook-nagradza-nienawisc-mark-zuckerberg-nie-chce-zmian-boi-sie-odwrotu-uzytownikow>.

⁵⁷ J. Kreft, *Władza platform*, dz. cyt., s. 250.

⁵⁸ *Sztuczna inteligencja walczy z mową nienawiści*, <https://www.uw.edu.pl/sztuczna-inteligencja-walczy-z-mowa-nienawisci> (20.09.2023).

zachowań użytkowników, jak i analizę reprezentacji ocenianych treści⁵⁹. Jest on świadomy, że agresja słowna jest zjawiskiem subiektywnym i że nowy pomysł może zamknąć ludzi w bańkach informacyjnych. Inna z kolei inicjatywa, prof. Katarzyny Budzyńskiej z Politechniki Warszawskiej, ma być krokiem w kierunku odzyskiwania zaufania do mediów społecznościowych. Jednym z wielu celów projektu przez nią kierowanego jest pomoc sztucznej inteligencji w działaniach interwencyjnych przeciwko mowie nienawiści, dezinformacji i polaryzacji⁶⁰. Czas pokaże efekty tych prac.

Drugim problemem hamującym skuteczność walki z mową nienawiści jest kwestia istnienia wielu grup zainteresowanych wzbudzaniem konfliktów, kłóceniem ludzi; ponadto najbardziej cyniczni wykorzystują media społecznościowe do wojny hybrydowej i sączenia nienawiści. Miliardy użytkowników wyrażają swoje opinie, dzielą się poglądami, reagują emocjonalnie na fakty dziejące się na ich własnym lub światowym podwórku, nie będąc świadomymi efektu komory echa⁶¹. Zdaniem Nowaka w mediach społecznościowych zachodzi niespotykany wcześniej proces oddolnego rozpowszechniania informacji, które, jeśli zostają uznane za ciekawe lub zabawne, są obecne w sieci na długo⁶². Obok treści rozrywkowych dotyczy to również treści kontrowersyjnych. Eksperti diagnozują epidemię nienawiści, na którą nie ma idealnej szczepionki. Internet działa wiralowo, rozsyłane treści są powielane i rozsyłane dalej, nawet gdy zostanie zlikwidowane ich źródło. Zdaniem Rogera McNamee narzędzia technologiczne „przy zaangażowaniu sztucznej inteligencji i automatyzacji stają się podatne na manipulacje”⁶³. Media społecznościowe bywają bezradne wobec ataków hejterskich i plagi fake newsów, często przygotowanych przez algorytmy krajów lub ośrodków skupiających władzę, celowo wywołujących elektroniczne wojny. Dochodzi do tego silna potrzeba kształtowania swojej tożsamości w oparciu o więzi

⁵⁹ L. Tomala, *Badacze pokazują, jak nauczyć sztuczną inteligencję wrażliwości użytkownika*, <https://naukawpolsce.pl/aktualnosci/news%2C96088%2Cbadacze-pokazuja-jak-nauczyc-sztuczna-inteligencje-wrazliwosci-uzytownika> (30.09.2023).

⁶⁰ *PW na czele projektu badającego mowę nienawiści i fake newsy*, <https://www.pw.edu.pl/aktualnosci/pw-na-czele-projektu-badajacego-mowe-nienawisci-i-fake-newsy> (8.10.2023).

⁶¹ A. Kaźmierska, W. Brzeziński, *Strefy cyberwojny*, Oficyna 4eM, Warszawa 2018, s. 94.

⁶² J. Nowak, dz. cyt., s. 20.

⁶³ R. McNamee, dz. cyt., s. 235.

społeczne milionów użytkowników internetu, wynikająca z lęku przed samotnością, pomijaniem i wykluczeniem. Ludzie, nie chcąc stracić relacji, koncentrują się na widoczności, traktując relacje w social mediach jako namiastkę bezpieczeństwa. Zdaniem Baumana i Lyona „użytkownicy Facebooka spowiadają się nieustannie w przenośnych elektronicznych konfesjonalach”⁶⁴. Algorytm lubi emocje, karmi się nimi i daje życie emocjonalnym wpisom.

Czy można bota nauczyć empatii i sprawiedliwości?

Po trzecie, największym problemem jest nauczenie sztucznej inteligencji retoryki nienawiści, słownictwa uważanego za obraźliwe, dyskryminujące, ksenofobiczne. Eksperti z branży IT twierdzą, że algorytmy potrzebują czasu i danych. O potencjale algorytmów Karen How mówi: „maszyna zanim zidentyfikuje mowę nienawiści, musi się jej nauczyć i być przeszkolona na nowych treściach”⁶⁵. Aby sztuczna inteligencja była w stanie zatrzymać wir wciągający użytkowników do wojenek słownych i wojen plemiennych, musi wiedzieć, czym jest to, przed czym powinna chronić. Ważkie pytanie wymaga odpowiedzi: jakie słowa i wyrażenia będą potraktowane przez platformy jako hejt? Pytania o wolność słowa przynoszą pytania o kulturę wypowiedzi, bo gdy jej brak, obrażani są przedstawiciele mniejszości, ale też chrześcijanie i ich symbole. Niestety, w samym środowisku polskich akademików odbywa się burzliwa dyskusja o istocie mowy nienawiści i nierozwiązany jest dylemat, czy obrażają tylko słowa, czy też symbole, czy może robią to wulgaryzmy, jak przedwyborcze osiem gwiazdek.

Kolejny dylemat istoty sztucznej inteligencji to fakt ujawniania polityczności decyzji osób zajmujących się algorytmami. W myślenie społeczne (o różnych grupach, ale i o swojej) jest wpisane kategoryzowanie, wyodrębnianie. Jaki wspólny mianownik wykorzysta administrator strony internetowej i czy są w ogóle poglądy neutralne, bez narosłych mentalnych klisz?

⁶⁴ Z. Bauman, D. Lyon, *Płynna inwigilacja. Rozmowy*. Wydawnictwo Literackie, Kraków 2013, za: M. Juza, *Między wolnością a nadzorem*, art. cyt., s. 255.

⁶⁵ J. Kreft, *Władza platform...*, dz. cyt., s. 111.

Decyzje, jak użyć sztucznej inteligencji do czyszczenia mowy nienawiści, będą nie tylko prawne i etyczne, lecz także polityczne. Nie te związane z konkretną partią, ale odpowiadające na pytanie o postawę wobec różnych niesprawiedliwości, praw mniejszości, emancypacji środowisk. Przecież maszyna zdecyduje tak, jak nauczył ją człowiek. Uczenie maszyny jest zawsze polityczne, zakłada określone przekonania. Krytycy mówią, że na tym innowacyjność maszyn się skończy, przyszłość bowiem będzie planowana na wzór przeszłości. Powielanie uprzedzeń widać było w Stanach Zjednoczonych, gdy działalność komisji ds. zwolnień warunkowych więźniów, stosujących dane sądów do przewidywania ryzyka recydywy, nosiła znamiona rasizmu⁶⁶. Bardzo prawdopodobne jest to, że problem eliminacji rasizmu ze świata on-line przyniesie spiętrzenia już obecnych problemów. Kultura kasowania, znana jako *cancel culture*, która z Ameryki trafiła już także do Polski, wymusi reakcję różnych środowisk na treści ich zdaniem obraźliwe. Dylemat kasować – nie kasować zastąpi szekspirowskie *to be or not be* i z pewnością zbuduje niejedną konflikt w mediach.

Odpowiedzialność platform i instytucji w spolaryzowanym świecie

Wielu autorów zgadza się z tezą, że liczne przestrzenie, których dotknie sztuczna inteligencja, powinny być uregulowane przez prawo, przy założeniu, że prawnicy będą rzecznikami cenionych powszechnie wartości, a nie adwokatami diabła, pracującymi na zlecenie tyranów lub socjopatów. Pojawia się pytanie o to, kto zajmie się selekcją prawników i kto ich uwrażliwi na sprawę tak delikatną, zniuansowaną, jaką jest język, wraz z jego frazeologią i słowotwórstwem. Czy firmy technologiczne będą dla prawników partnerem czy petentem? Facebook i Twitter są chronione ustawą *Communications Decency Act* z 1996 r., odpowiedzialność za publikację treści jest nałożona na internautów. Sprawę jeszcze bardziej komplikuje fakt, że w polskim prawie

⁶⁶ Tamże, s. 232.

karnym nie ma definicji mowy nienawiści i podczas procesów sądowych prokuratorzy wykorzystują art. 256 i 257 k.k. dla jej wąskiego lub szerokiego określenia, decydując o tym, czy jakieś działanie było przestępstwem z nienawiści lub czy jakaś wypowiedź miała charakter znieważający i poniżający (akt języka pogardy)⁶⁷.

Realia pokazują, że technologia wchodzi do społeczeństwa skłóconego, spolaryzowanego, zależnego od technologii mediów społecznościowych i w dużej mierze uzależnionego. Kluczowe będą decyzje i wybory dokonywane przez ludzi zajmujących się SI: właściciele firm technologicznych, usługobiorców, czyli tych, których stać będzie na zakup technologii czyszczącej internet, np. przywódców rządów zainteresowanych leczeniem internetu zainfekowanego wpisami trolli, i wreszcie informatyków pracujących z algorytmami, projektujących, co należy usuwać. Zdaniem T. Williamsa, badacza z Uniwersytetu Huddersfield, narażenie na mowę nienawiści i treści ekstremistyczne obecne w przestrzeni on-line prowadzi do „większego prawdopodobieństwa wystąpienia aktów przemocy”⁶⁸ w społeczeństwie. I choć nie ma dokładnych badań kryminologów na ten temat, korelacje między przemocą słowną a przemocą fizyczną są prawdopodobne.

Kolejne ważne pytanie na przyszłość dotyczy sposobu współpracy instytucji i organizacji w zarządzaniu sztuczną inteligencją. Która z nich: państwowa czy europejska, będzie miała ostatnie słowo w rozstrzygnięciu dylematów, co należy uznać za hejt i kogo obarczać sankcjami: użytkowników czy właściciele social mediów? Na polskim podwórku panuje rozdźwięk między stygmatyzacją mowy nienawiści a przyzwoleniem na nią w życiu politycznym. Rzecznik Spraw Obywatelskich w 2022 r. wyraził oburzenie działaniem Facebooka, który blokował lub usuwał profile w sposób niejasny, łamiąc wolność słowa zapisaną w konstytucji. Apelowal o tzw. weryfikację ręczną i kontekstową⁶⁹, a nie automatyczną, obwiniając platformę o zbytnią ufność w pracę algorytmów.

⁶⁷ M. Urbańczyk, *Definiowanie mowy nienawiści w orzecznictwie sądów polskich w sprawach karnych*, „Studia nad Autorytaryzmem i Totalitaryzmem” 2020, t. 42, nr 1, s. 82.

⁶⁸ *Media społecznościowe pełne mowy nienawiści a serwisy sobie z tym nie radzą. Pomoże sztuczna inteligencja*, <https://www.bankier.pl/wiadomosc/Media-spolescnosciowe-pelne-mowy-nienawisci-a-serwisy-sobie-z-tym-nie-radza-Pomoze-sztuczna-inteligencja-8640770.html> (11.10.2023).

⁶⁹ *Usuwanie komentarzy i blokowanie kont na Facebooku. Opinia Rzecznika*, 25.07.2022, <https://bip.brpo.gov.pl/pl/content/rpo-facebook-usuwanie-komentarzy-blokowanie-kont>.

Mowa nienawiści w internecie – social media między funkcją misyjną a biznesową

Uproszczeniem tematu jest teza, że sztuczna inteligencja wyręczy człowieka i wyczyści bałagan oraz gruzy po kłótniach, jakie zostawiamy w internecie, ślady zostawiane przez polityków lub złośliwców szydzących z sukcesów innych. Fora internetowe, czasopisma, platformy opiniotwórcze, wszyscy usługodawcy tworzący dziś elektroniczną agorę, chcąc poradzić sobie z negatywnymi komentarzami internautów, zapytają, czy przesianie negatywnych wpisów jest możliwe za pomocą ludzkich rąk – przecież internet jest elektroniczną agorą i kakofonią bodźców i słów.

Wielcy gracze na rynku usług technologicznych (Google, Microsoft, IBM, Facebook, Twitter) są postrzegani jako podmioty innowacyjne, od których oczekuje się bezpiecznych usług internetowych. Jednak oni sami poprawnymi i świętymi być nie chcą, przepraszają za wpadki tylko pod presją społeczną i wskutek znaczącego odpływu użytkowników. Używają programów segregujących wiadomości w różnych celach, tak dla wzmocnienia zaangażowania i uzależnienia od oglądania, jak i dla selekcji nienawistnych treści; algorytm to dla nich miecz obosieczny – takiej metafory użył niedawno sam Elon Musk⁷⁰. Badacz kultury organizacyjnej i zarządzania Jan Kreft przypomina, że platformy wyrosłe w Dolinie Krzemowej, takie jak Facebook i Twitter, są przeniknięte duchem matematyczno-inżynierskim i pragmatyzmem, a rozwiązywanie problemów jest skoncentrowane na tworzeniu algorytmów, a więc zapisu nowych kodów, ciągu liczb. Jego zdaniem nie należy spodziewać się po nich działania etycznego, gdyż mają one w sobie zakodowane wartości rynkowe, a nakłanianie do korzystania z produktów lub usług służy celom promarketingowym⁷¹. Eksperti twierdzą, że firmy technologiczne żyją w atmosferze wielkiej rywalizacji, zajęte wyniszczaniem konkurentów i zdobywaniem nowych rynków. Dominuje w nich logika korporacyjnej efektywności. Czy mając nadzieję na sprzątanie przez

⁷⁰ *Tech titans meet US lawmakers, Musk seeks 'referee' for AI*, Reuters, 14.09.2023, <https://www.reuters.com/technology/musk-zuckerberg-gates-join-us-senators-ai-forum-2023-09-13/> (8.01.2024).

⁷¹ J. Kreft, *Władza algorytmów. U źródeł potęgi Google i Facebooka*, Wydawnictwo UJ, Kraków 2019, s. 47.

nie internetu z obraźliwych wpisów, nie wyrażamy bezkrytycznej afirmacji technologii? Facebook i Twitter chcą rozszerzać imperium wpływu i być innowacyjne w tym, co robią. Nadzieję niosą organizacje czuwające nad bezpieczeństwem użytkowników sieci i różnego pochodzenia organizacje obywatelskie (Responsible AI Institute, grupa robocza OECD ds. AI⁷²), dbające o to, by algorytmy nie były stronnicze. Nadzieję może nieść konkurencja między platformami w temacie odpowiedzialności, czyli wyścig o to, która firma będzie bardziej wolna od hejtu i bardziej odpowiedzialna społecznie, jednak czy ta kwestia jest przez właścicieli owych platform pożądana, stanowi tymczasem tylko pytanie akademickie.

Sztuczna inteligencja z pewnością nie może być jedynym narzędziem rozwiązywania problemów. Technologia to zautomatyzowane narzędzia, nie mogą być wyłącznym sposobem na trzymanie w ryzach tendencji do agresywności; nie ochroni też ludzi przed nimi samymi. Stworzy jedynie lepszy algorytm. Obietnice start-upów wydają się i obiecujące, i ryzykowne. Czy firmy technologiczne mają sprzątać bałagan w internecie? Mogłyby przecież zająć się wieloma pożytecznymi sprawami. A może chodzi im o wcielanie w życie „nowej ewangelii”, nowej „technoutopii”⁷³ obiecującej naprawienie świata? Sztuczna inteligencja nie ma ani krzty inteligencji, ani krzty inwencji⁷⁴, mówi prof. Kisielewicz, a więc będzie pracować jak typowa maszyna, według zapisanych w niej kodów. Będzie uczona ludzkich sposobów myślenia, bo inżynierowie sztucznej inteligencji będą pracować nad modelami językowymi kopiującymi ich (lub ich zleceńodawców) społeczne postrzeganie świata.

Algorytmy sztucznej inteligencji w postulatach ekspertów mają pomóc w dezynfekcji internetu, skażonego wirusem negatywnych emocji. Obok mesjańskiej i funkcjonalno-ekonomicznej roli technologii coraz częściej pojawia się w społeczeństwie marzenie, by uporządkowała ona nasz świat. Nie uporządkuje. Paradoksalne jest i to, że algorytmy walczą z trudnymi,

⁷² *Sztuczna inteligencja jak Internet w 1995 roku*, <https://cyfrowa.rp.pl/opinie-i-komentarze/art39248221-sztuczna-inteligencja-jak-internet-w-1995-roku-nadchodzi-niezwykly-czas> (12.10.2023).

⁷³ J. Kreft, *Władza platform...*, dz. cyt., s. 249.

⁷⁴ A. Kisielewicz, *Bajki o sztucznej inteligencji i prawdziwe zagrożenia*, <https://wszystkoconajwazniejsze.pl/andrzej-kisielewicz-bajki-o-sztucznej-inteligencji-i-prawdziwe-zagrozenia/> (6.10.2023).

bardzo silnymi emocjami, które powinny być zarządzane przez psycho-terapeutów. Węzeł gordyjski w pytaniu o dialog człowieka z technologią zacieśnia się w obliczu coraz to nowych konfliktów społeczno-politycznych. Współczesne społeczeństwo oczekuje od sztucznej inteligencji reagowania na agresję słowną stworzoną przez człowieka. Czy sztuczna inteligencja, fetyszyzowana i idealizowana, zatrzyma w internecie powódź komentarzy, w których hejtujemy, bo z różnych powodów nie umiemy rozmawiać? Czy robota można nauczyć przebaczać, reagować dobrem na zło, wyciszać spory i kłótnie, których nie umie wyciszać sam człowiek? Tego oczekują nauka i biznes. Robot wprowadzi porządku pozorne, bo z trudnymi emocjami i problemami z komunikacją człowiek powinien (musi) radzić sobie sam.

Bibliografia

- Amerykańskie media alarmują: Facebook przyczynił się do „społecznej wojny domowej” w Polsce*, 30.10.2021, <https://www.polskieradio.pl/129/1280/Artykul/2837300%-2CAmerykanskie-media-alarmuja-Facebook-przyczynil-sie-do-spoecznej-wojny-domowej-w-Polsce>, Polskie Radio (30.08.2022).
- Castells M., *Władza komunikacji*, PWN, Warszawa 2013.
- Definicja Sztucznej Inteligencji (SI), [https://www.europarl.europa.eu/meet docs/2014_2019/plmrep/COMMITTEES/JURI/DV/2019/11-06/Ethics-guidelines-AI_PL.pdf](https://www.europarl.europa.eu/meet/docs/2014_2019/plmrep/COMMITTEES/JURI/DV/2019/11-06/Ethics-guidelines-AI_PL.pdf) (12.10.2023).
- Disney, IBM, Apple i inne marki wycofują reklamy z X/Twittera. Powód: antysemityczne treści*, <https://www.wirtualnemedial.pl/artykul/koniec-reklam-x-twitter-disney-ibm-apple-antysemityzm> (30.08.2023).
- Facebook and YouTube moderators sign PTSD disclosure*, BBC, 25.01.2020, <https://www.bbc.com/news/technology-51245616> (12.10.2023).
- Facebook pozwał na brutalne nienawistne wpisy? Naukowcy i aktywiści chcą 2 mld dol. dla ofiar przemocy*, 15.12.2022, <https://www.wirtualnemedial.pl/artykul/facebook-pozwany-mowa-nienawisci> (12.10.2023).
- Facebook zapłaci 52 mln dolarów odszkodowania moderatorom za pracę w trudnych warunkach*, <https://www.wirtualnemedial.pl/artykul/facebook-zaplaci-52-mln-dolarow-odszkodowania-moderatorom-za-prace-w-trudnych-warunkach> (13.10.2023).
- Flores S.E., *Sfejsowani. Jak media społecznościowe wpływają na nasze życie, emocje i relacje z innymi*, Muza, Warszawa 2017.
- Forst K., *Facebook we współczesnym świecie komunikacji społecznej*, w: *Media społecznościowe: dialog w cyberprzestrzeni*, red. K. Doktorowicz, t. 2, Wydawnictwo Uniwersytetu Śląskiego, Katowice 2017, s. 37–55.

- Fraser M., *Moderacja treści na Facebooku oparta o sztuczną inteligencję nie działa*, 18.10.2021, <https://cyberdefence24.pl/social-media/moderacja-tresci-na-facebooku-oparta-o-sztuczna-inteligencje-nie-dziala> (30.09.2023).
- Frącz A., *Hejting – przejawy mowy nienawiści w Sieci*, „Parezja” 2019, nr 2, s. 29–47.
- Frenkel Sh., Kang C., *Brzydka prawda. Kulisy walki Facebooka o dominację*, Rebis, Poznań 2022.
- Galloway S., *Wielka czwórka. Four. Ukryte DNA: Amazon, Apple, Facebook i Google*, Rebis, Poznań 2018.
- Głowiński M., *Retoryka nienawiści*, „Nauka” 2007, nr 2, s. 19–27.
- Goczał B., *Algorytm Facebooka nagradza nienawiść*, 10.10.2022, <https://www.wirtualnemedia.pl/artukul/algorytm-facebook-nagradza-nienawisc-mark-zuckerberg-nie-chce-zmian-boi-sie-odwrotu-uzytkownikow> (30.09.2023).
- Hejterstwo. Nowa praktyka kulturowa? Geneza, przypadki, diagnozy*, red. J. Dynkowska, N. Leman, M. Wróblewski, A. Zatora, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2017.
- Internet. Platformy odpuszczają usuwanie mowy nienawiści*, DW, 26.11.2022, <https://www.dw.com/pl/internet-platformy-odpuszczaj%C4%85-usuwanie-mowy-nienawiw%C5%9Bci/a-63899752> (30.09.2023).
- Internetowi potentaci razem przeciw mowie nienawiści, Facebookowi pomoże w tym sztuczna inteligencja*, <https://www.dobreprogramy.pl/internetowi-potentaci-razem-przeciw-mowie-nienawisci-facebookowi-pomoze-w-tym-sztuczna-inteligencja,6628359852615809a> (30.09.2023).
- Iwanowska M., *Zrozumieć media (społecznościowe). Perspektywa psychologiczna*, w: *Media, biznes, kultura. Rzeczywistość medialna. Formy, problemy, aspiracje*, red. M. Łosiewicz, A. Ryłko-Kurpiewska, Novae Res, Gdynia 2016, s. 169–185.
- Juza M., *Hejterstwo w komunikacji internetowej: charakterystyka zjawiska, przyczyny i sposoby przeciwdziałania*, „Profilaktyka Społeczna i Resocjalizacja” 2015, nr 25, s. 27–50.
- Juza M., *Między wolnością a nadzorem. Internet w zmieniającym się społeczeństwie*, Wydawnictwo Naukowe Scholar, Warszawa 2019.
- Kaźmierska A., Brzeziński W., *Strefy cyberwojny*, Oficyna 4eM, Warszawa 2018.
- Kisielewicz A., *Bajki o sztucznej inteligencji i prawdziwe zagrożenia*, Wszystko co Najważniejsze, <https://wszystkocoonajwazniejsze.pl/andrzej-kisielewicz-bajki-o-sztucznej-inteligencji-i-prawdziwe-zagrozenia/> (6.10.2023).
- Komisja Europejska zastrza zasady moderacji treści*, 27.04.2023, <https://www.wirtualnemedia.pl/artukul/komisja-europejska-zastrza-zasady-moderacja-tresci-facebook-tiktok-twitter-google-wikipedia-apple> (11.10.2023).
- Krawczyński J., *Rząd USA nałożył 5 mld kary dla Facebooka*, 14.07.2019, <https://www.purepc.pl/trzad-usa-nalozyl-5-mld-kary-dla-facebooku-firma-na-tym-zyskala> (10.12.2023).
- Kreft J., *Władza algorytmów. U źródeł potęgi Google i Facebooka*, Wydawnictwo UJ, Kraków 2019.
- Kreft J., *Władza platform. Za fasadą Google, Facebooka i Spotify*, Universitas, Kraków 2022.
- Lee K.-F., *Inteligencja sztuczna, rewolucja prawdziwa. Chiny, USA i przyszłość świata*, tłum. K. Hejwowski, Media Rodzina, Poznań 2019.

- Leszczuk-Fiedziukiewicz A., *Czy hejt i mowa nienawiści staną się normą? Społeczne uwarunkowania zachowań dewiacyjnych w Internecie*, „Media – Kultura – Komunikacja Społeczna” 2018 nr 14/3, s. 99–117.
- Leszczuk-Fiedziukiewicz A., *Inicjatywy antyhejtowe. Wybrane przykłady zaangażowania społecznego w ograniczanie mowy nienawiści*, w: *Język a media. Perspektywy i zagrożenia języka we współczesnych mediach*, red. E. Horyń, B. Skowronek, A. Walecka-Rynduch, Collegium Columbianum, Kraków 2020, s. 99–115.
- Levy S., *Facebook. A miało być tak pięknie. Za kulisami imperium z Doliny Krzemowej*, tłum. K. Sosnowska, Białystok 2022.
- McNamee R., *Nabici w Facebooka. Przestroga przed katastrofą*, tłum. K. Puławski, Media Rodzina, Poznań 2019.
- Media społecznościowe pełne mowy nienawiści, a serwisy sobie z tym nie radzą. Pomoże sztuczna inteligencja*, <https://www.bankier.pl/wiadomosc/Media-spolesnosciove-pejne-mowy-nienawisci-a-serwisy-sobie-z-tym-nie-radza-Pomoze-sztuczna-inteligencja-8640770.html> (11.10.2023).
- Mowa nienawiści na Twitterze. Sztuczna inteligencja ma wspierać moderację*, <https://cyberdefence24.pl/social-media/mowa-nienawisci-na-twitterze-sztuczna-inteligencja-ma-wspierac-moderacje> (30.08.2023).
- Naruszewicz-Duchlińska A., *Nienawiść w czasach Internetu*, Novae Res, Olsztyn 2015.
- Nijakowski L., *Mowa nienawiści w świetle teorii dyskursu*, w: A. Horolets, *Analiza dyskursu w socjologii i dla socjologii*, Adam Marszałek, Toruń 2008, s. 113–133.
- Nowak J., *Social media jako sieci obiegu przekazów medialnych*, „Nowe Media” 2012, nr 3, s. 13–31.
- Nowina Konopka M., „Każdy zna się na AI”. *Przegląd badań polskiej opinii publicznej na temat sztucznej inteligencji*, „Zeszyty Prasoznawcze” 2023, t. 66, nr 4 (256), s. 127–142.
- Orzeł W., *Czy sztuczna inteligencja poradzi sobie z mową nienawiści?*, <https://aibusiness.pl/czy-sztuczna-inteligencja-poradzi-sobie-z-mowa-nienawisci/> (10.10.2023).
- Osowiecka M., *Nienawistny język w Internecie*, „Ja. My. Oni. Poradnik Psychologiczny Polityki”, 26.04.2015, t. 18, <https://www.polityka.pl/jamyoni/1617093,1,nienawistny-jezyk-w-internecie.read> (11.10.2023).
- PW na czele projektu badającego mowę nienawiści i fake newsy*, <https://www.pw.edu.pl/aktualnosci/pw-na-czele-projektu-badajacego-mowe-nienawisci-i-fake-newsy> (8.10.2023).
- Pyżalski J., *Agresja elektroniczna i cyberbullying jako nowe ryzykowne zachowania młodzieży*, Impuls, Kraków 2012.
- Simons J., Frankel E., *Dlaczego sztuczna inteligencja powinna być demokratyczna?*, Wszystko co Najważniejsze, 17.03.2023, <https://wszystkoconajwazniejsze.pl/josh-simons-i-eli-frankel-sztuczna-inteligencja-demokratyczna/> (10.08.2023).
- Sztuczna inteligencja będzie cenzorem na FB*, „Frona”, 19.05.2020, <https://www.frona.pl/a/dzieki-nowym-rozwiazaniom-si-cenzura-na-facebooku-bedzie-skuteczniejsza,144531.html> (2.10.2023).
- Sztuczna inteligencja jak Internet w 1995 roku. „Nadchodzi niezwykły czas”*, <https://cyfrowa.rp.pl/opinie-i-komentarze/art39248221-sztuczna-inteligencja-jak-internet-w-1995-roku-nadchodzi-niezwykly-czas> (12.10.2023).

- Sztuczna inteligencja walczy z mową nienawiści*, <https://www.uw.edu.pl/sztuczna-inteligencja-walczy-z-mowa-nienawisci>, <https://www.uw.edu.pl/sztuczna-inteligencja-walczy-z-mowa-nienawisci/> (20.09.2023).
- Tech titans meet US lawmakers, Musk seeks 'referee' for AI*, Reuters, 14.09.2023, <https://www.reuters.com/technology/musk-zuckerberg-gates-join-us-senators-ai-forum-2023-09-13/> (8.01.2024).
- Tomala L., *Badacze pokazują, jak nauczyć sztuczną inteligencję wrażliwości użytkownika*, <https://naukawpolsce.pl/aktualnosci/news%2C96088%2Cbadacze-pokazuja-jak-nauczyc-sztuczna-inteligencje-wrazliwosci-uzytkownika> (20.09.2023).
- Traumatized Facebook moderators turn to sex on the job*, „New York Post”, 25.02.2019, <https://nypost.com/2019/02/25/trauma-a-job-hazard-for-facebook-content-moderators/> (12.10.2023).
- Twitter usunął 70 mln fałszywych kont*, 10.07.2018, <https://www.euractiv.pl/section/gospodarka/news/twitter-usunal-70-mln-falszywych-kont-w-dwa-miesiace/> (10.08.2023).
- Tymieniecka-Cichoń A., *Mowa nienawiści a media społecznościowe*, w: *Mowa nienawiści a prawo na tle współczesnych zjawisk społeczno-politycznych*, red. W. Pływaczewski, M. Duda, Uniwersytet Warmińsko-Mazurski, Olsztyn 2017, s. 156–162.
- Urbańczyk M., *Definiowanie mowy nienawiści w orzecznictwie sądów polskich w sprawach karnych*, „Studia nad Autorytaryzmem i Totalitaryzmem” 2020, 42, nr 1, s. 79–100.
- Usuwanie komentarzy i blokowanie kont na Facebooku. Opinia Rzecznika*, 25.07.2022, <https://bip.brpo.gov.pl/pl/content/rpo-facebook-usuwanie-komentarzy-blokowanie-kont> (20.09.2023).
- Vaidhyanathan S., *Antisocial Media. Jak Facebook oddala nas od siebie i zagraża demokracji*, Wydawnictwo W.A.B, Warszawa 2018.
- Więcej niż polityka*, Michał Bilewicz w rozmowie z Danielem Lisem, „Znak” 2019, nr 4, s. 22–27.
- Winczewska-Latoszek K., *Wyzwania i zagrożenia sztucznej inteligencji*, <https://web.swps.pl/strefa-zarzadzania/artykuly/19084-to-ja-robot-wyzwania-i-zagrozenia-sztucznej-inteligencji?dt=1704380584740> (13.10.2023).
- Żardecka-Nowak M., *Resentyment jako postać nienawiści*, w: *Nienawiść w życiu publicznym*, red. W. Zuziak, J. Synowiec, Wydawnictwo Naukowe Uniwersytetu Papieskiego Jana Pawła II w Krakowie, Kraków 2016.
- Żmuda B., *Czy może istnieć nienawiść usprawiedliwiona? Problem nienawiści ofiar do oprawców w myśli filozoficznej Barbary Skargi*, w: *Nienawiść w życiu publicznym*, red. W. Zuziak, J. Synowiec, Wydawnictwo Naukowe Uniwersytetu Papieskiego Jana Pawła II w Krakowie, Kraków 2016.

The role of artificial intelligence in the fight against hate speech in social media on the example of Facebook and Twitter

Summary

The purpose of the article is to attempt to analyze the policies of social media owners towards the issue of using artificial intelligence, using the example of two technology giants Facebook and Twitter. Their demands, initiatives and numerous inconsistencies in the treatment of so-called harmful content, perceived as hate speech, will be discussed. The researcher uses the method of analyzing found data, using publications that expose the behind-the-scenes work of Facebook, media discourse analysis focused on expert evaluation of the monopolists' actions. The article's conclusions focus on the complex role that artificial intelligence plays in content and communications management, as Facebook and Twitter assure the public that they are using algorithms to eliminate racist, offensive material while pursuing their business and marketing goals. The algorithms they use act as a double-edged sword, driving traffic and engagement online and scouring the Internet of outrageous posts. It will be critical for public pressure and cooperation with institutions to ensure that the machine learning used to remove hate speech is ethical and responsible.

Keywords: hate speech, hate, social media, artificial intelligence (AI), machine learning, social media responsibility

ChatGPT, BARD, ERNIE – czy algorytmy sztucznej inteligencji to autorytety cyfrowego jutra?

Socjologiczne próby opisanie i interpretacji zmian zachodzących aktualnie w otaczającej nas rzeczywistości są zadaniem trudnym, zmuszają bowiem autora gotowego podjąć się wyzwania do zmierzenia się z teraźniejszością przesyconą wydarzeniami i informacjami. Trudność ta wynika też z faktu, że krajobrazy za oknem, przez które ludzie patrzą na świat, odpowiadają heraklitejskiej wizji ciągłej przemiany, a świat, który jest stale w ruchu, zmaga się z trwałym deficytem ważnego składnika – czasu, niezbędnego do spojrzenia z dystansu, a więc i możliwości refleksji. Czas – jak pisze Barbara Skarga „(...) przynosi selekcję wydarzeń i zjawisk, uwypukla całości, wyznacza punkty przecięć, wskazuje swym rytmem na okresy względnej stabilności bądź rewolucyjnego przyspieszenia”².

Tymczasem w ciągu zaledwie ostatnich kilku lat doświadczyliśmy wojny i pandemii, zła i lęku, dwóch ekstremalnych żywiołów, które w mniej lub bardziej dotkliwy sposób odcisnęły swoje piętno, narzucając bezprecedensowe zmiany w każdym aspekcie życia, w tym międzyludzkich relacji, nieodwracalnie skazanych na funkcjonowanie dwutorowe – w przestrzeniach

¹ Dr Anna Turner, Zakład Teorii Kultury, Instytut Filozofii i Socjologii, Polska Akademia Nauk.

² B. Skarga, *Granice historyczności*, PIW, Warszawa 1989, s. 9.

offline i online. Covidowy lockdown i przymus pozostania w domu doprowadziły do nagłego zderzenia dwóch przeciwstawnych sił, świata fizycznego i świata wirtualnego, być może na zawsze, we wszystkim, co robimy, odciskając widoczny cień tej dwubiegunowości. A przecież życie społeczne nie może toczyć się w sposób niepewny, nieuporządkowany, chaotyczny. Konieczne są reguły i normy kulturowe, przez stulecia kreowane przez jasno określone grupy wpływu: Kościół (za pomocą Dekalogu) i elity intelektualne (za pomocą kodeksów moralnych). Tutaj nastąpiła wyraźna zmiana. Rozwój mediów, a także coraz większe zakorzenienie społeczeństw w świecie cyfrowym doprowadziły do fragmentaryzacji widoczności tych grup, a w rezultacie do erozji ich sprawczości i zmniejszenia wpływu na funkcjonowanie społeczeństwa. To internet stał się drogowskazem, zaufanym miejscem, w którym szukamy i znajdujemy odpowiedzi na wiele nurtujących nas pytań.

Wobec takich komplikacji społecznej świadomości, w poszukiwaniu równowagi między tradycją a postępem, między wolnością a bezpieczeństwem, między człowiekiem a technologią, potrzebujemy prawdziwych autorytetów, tym bardziej że nasze lęki i niepewności to wdzięczne paliwo dla firm i polityków budujących na nich swój kapitał. W erze cyfrowej postprawdy i dezinformacji zyskuje na aktualności diagnoza Ulricha Becka:

Większość współczesnych zagrożeń nie poddaje się osobistej analizie i nie można ich wiarygodnie potwierdzić bądź zakwestionować na własną rękę, więc równie łatwo je publiczności „wmówić”, jak ją od nich «odwieść». A w starciu opinii największe szanse na zwycięstwo mają posiadacze największych sił medialnych³.

W wyniku rozwoju nowych technologii, w tym wprowadzenia na rynek niedrogich smartfonów z dostępem do internetu, odeszliśmy od uproszczonego, linearnego rozumienia świata, w którym *a* powoduje *b*, co z kolei powoduje *c*, do bardziej złożonego, wielowymiarowego zrozumienia gdzie *a* powoduje *b* w kontekście alfa i *a* powoduje *c* w kontekście beta. To istotne, ponieważ podejmując się udziału w tej dyskusji, chcę zwrócić uwagę na aspekty, które

³ Za: Z. Bauman, *Płynny lęk*, Wydawnictwo Literackie, Kraków 2008, s. 255.

splatają i przenikają się nawzajem. Zacznę więc od zmian, które zachodzą w środowisku naukowym, ze względu na ich duży wpływ na funkcjonowanie kultury eksperckiej i w rezultacie na stopień ich adaptacji do otaczającej nas cyfrowej rzeczywistości. Następnie przedstawię, jak kultura internetu doprowadziła do przeobrażeń w sferze komunikacyjnej, polegających na zastąpieniu pasywnego przekazu i odbioru wiedzy aktywnym poszukiwaniem informacji, a także jakie w związku z tym powstały nowe sfery wykluczeń. Przyjrzę się również najnowszym wydarzeniom rewolucjonizującym definicję eksperta, ponieważ do zmagania o to miano giganci świata cyfry wystawili algorytmy sztucznej inteligencji, które stają w szranki nie z ludźmi, lecz ze sobą.

Ewolucja kultury eksperckiej

Koncepcja kultury eksperckiej odnosi się do wspólnych wartości, norm, przekonań i praktyk, które charakteryzują jednostki mające specjalistyczną wiedzę i umiejętności w określonej dziedzinie. Aby nadać humanistyczny sens moim rozważaniom, proponuję zacząć od spojrzenia na kwestie kultury eksperckiej i tworzących ją autorytetów w kontekście współczesnych nauk społecznych oraz coraz wężziej definiowanych dyscyplin naukowych.

Szymon Wróbel, filozof, w swoim eseju o interdyscyplinarności pisze:

Po rozbiórce filozofii jako dyscypliny scalającej ludzką wiedzę i zdyskredytowaniu wiedzy potocznej, nauka wybrała ścieżkę wąskich specjalizacji. Poznanie naukowe – w rezultacie – dało się poznać nie poprzez bezgraniczność, ale właśnie graniczność, nie poprzez interdyscyplinarność, ale poprzez dyscyplinarność. (...) jedyną osobą naprawdę i autentycznie interdyscyplinarną był Arystoteles (...), mógł z powodzeniem oddawać się badaniom fizycznym, matematycznym, logicznym, anatomicznym, botanicznym, politologicznym, etycznym, ekonomicznym, wreszcie psychologicznym i metafizycznym, albowiem nie doświadczał jeszcze kategorycznego podziału fakultetów w akademii, której będzie już doświadczał Leonardo da Vinci. Da Vinci (...) przekraczał co prawda (...) granicę między nauką i sztuką, (...) przetrwa jednak raczej jako malarz i architekt, natomiast mniej jako inżynier i filozof. Jego rozważania na

temat maszyn wojennych, ruchów wody, anatomii, flory, fauny, aerodynamiki są bardziej świadectwem jego nieskrępowanej wyobraźni niż rzeczywistej dyscypliny naukowej. (...) ryzyko interdyscyplinarności będzie się dokonywało od czasów da Vinci pod coraz większym zagrożeniem ignorancją. (...) Stało się tak dlatego, że da Vinci, w przeciwieństwie do Arystotelesa, nie dysponował już wiedzą o całości, nie posiadał już całościowego modelu uprawiania wiedzy, który porządkowałby i scalał jego cząstkowe poczynania poznawcze⁴.

Zatrzymajmy się nad tym na chwilę. Jest faktem, że w toku nieustannego rozwoju nauki musi dochodzić do jej systematyzacji przez podziały na kategorie, dziedziny, dyscypliny i fakultety, tym samym mamy do czynienia z sytuacją, w której reprezentujący je eksperci specjalizują się w wąskich obszarach badawczych, często rozumiały tylko dla hermetycznego grona badaczy zajmujących się podobnymi zagadnieniami. Skutkiem ubocznym takiego rozproszenia wiedzy, kiedyś skupionej wokół zaledwie kilku głównych źródeł, jest to, że staje się ona nieprzystępna dla laików, którzy, kierowani ludzkim dążeniem do poznania i zrozumienia rzeczywistości, przeniosą swoje poszukiwania w inne, bardziej im przyjazne miejsca, także cyfrowe, doprowadzając do sytuacji, w której coraz więcej doświadczeń i interakcji będzie pochodziło od technologii, a nie od ludzi. To rzecz warta odnotowania. Trzeba zdać sobie sprawę, na czym ten proces polega i do czego może doprowadzić, szczególnie w rzeczywistości, która stając się coraz bardziej „smart”, skraca dystans między pragnieniem a jego realizacją, i już tylko kliknięcie dzieli użytkownika od znalezienia odpowiedzi na wielkie pytania: kim jestem, dokąd zmierzam, co chcę robić.

Nauka ściśle zarezerwowana dla lepiej wykształconych i całkowicie niedostępna dla pewnych grup społecznych jest nauką, która wyklucza, wbrew własnym interesom. Dopóki pozostanie zbyt elitarna, funkcjonująca w oderwaniu od codzienności zwykłego człowieka, będzie to proces nieodwracalny. Nauka ma obowiązek służyć całej populacji, a wykluczenie niektórych grup może utrzymywać stereotypy i wzmacniać uprzedzenia, utrudniając

⁴ Sz. Wróbel, *Interdyscyplinarność jako efekt dyscyplinarności*, w: *Głosy w sprawie interdyscyplinarności*, red. J. Kurczewska, M. Lejzerowicz, Wydawnictwo IFiS PAN, Warszawa 2014, s. 29–30.

postęp w zrozumieniu i rozwiązywaniu kwestii, które nieproporcjonalnie wpływają na zmarginalizowane społeczności. Uważam, że to różnorodność jest katalizatorem postępu. Po pierwsze, różnorodność promuje wielość perspektyw i pomysłów, wspierając innowacyjność i kreatywność. Po drugie, jest niezbędna do skutecznego radzenia sobie z wyzwaniami globalnymi. Złożone kwestie, takie jak zmiany klimatu, kryzysy zdrowia publicznego i rozwój technologiczny, wymagają współpracy interdyscyplinarnej. Uwzględniając różnorodne głosy i wiedzę specjalistyczną, społeczność naukowa zostaje lepiej przygotowana do zrozumienia wieloaspektowego charakteru tych wyzwań i opracowania kompleksowych, zrównoważonych rozwiązań. Włączenie ludzi z różnych środowisk pozwoli zobaczyć, że są oni reprezentowani w nauce, co pomoże przełamać bariery i budować zaufanie wśród społeczeństwa, ponieważ umożliwia zaangażowanie się w pracę naukową osobom z niedostatecznie reprezentowanych grup społecznych.

Kultura internetu – rola widoczności

W kontekście przeobrażeń, którym poświęcam uwagę w tym artykule, należy wspomnieć o istotnych wydarzeniach zaistniałych w ciągu zaledwie dekady. To dwie przełomowe zmiany, dwie rewolucyjne innowacje. Pierwsza z nich to pojawienie się smartfonów, a więc telefonów z aparatem cyfrowym i dostępem do internetu. To małe urządzenie, które McLuhan nazwałby *przedłużeniem człowieka*, dało nam dwie możliwości: po pierwsze, rejestrację i utrwalanie tego, co dzieje się wokół nas; po drugie, możliwość dzielenia się tym z innymi i wchodzenie ze sobą w interakcje, które ułatwiły kontakty z bliskimi, jednocześnie przenosząc aktywność społeczną w sferę przestrzeni internetowej. Nie byłoby to jednak możliwe, gdyby nie pojawienie się platform internetowych, takich jak media społecznościowe, które zmieniły społeczeństwo na wiele sposobów, tym samym nieodwracalnie przekształcając komunikację, rozpowszechnianie informacji i interakcje społeczne. Platformy cyfrowe, takie jak Facebook, Instagram, TikTok czy Twitter, a także komunikatory, takie jak WhatsApp, Messenger, Zoom czy Skype, umożliwiły nam dzielenie się swoimi przemyśleniami i doświadczeniami w czasie rzeczywistym, rzucając tym samym wyzwanie mediom tradycyjnym i na zawsze zmieniając

sposób, w jaki wchodzimy w interakcje z ludźmi i technologią. Odtąd w bardzo prosty sposób możemy relacjonować światu, gdzie jesteśmy, co widzimy, czego słuchamy, a dla wielu nic nie dzieje się naprawdę, dopóki nie zostanie utrwalone i pokazane w cyfrowym, lustrzanym odbiciu.

Takie wzajemne powiązania mają zarówno pozytywne, jak i negatywne konsekwencje. Z jednej strony skracają dystans, wspierając poczucie bycia mieszkańcem *globalnej wioski*, a z drugiej przyczyniają się do błyskawicznego rozprzestrzeniania się dezinformacji, co znacząco wpłynęło na krajobraz polityczny, czyniąc z mediów społecznościowych potężne narzędzie mobilizacji politycznej i wywierania wpływu. Skutki tego mogliśmy obserwować w inicjatywach takich jak arabska wiosna czy #BlackLivesMatter, a także w wyborach prezydenckich czy w kampanii, która w Wielkiej Brytanii doprowadziła do brexitu. Jednocześnie na zawsze zmieniły się krajobrazy prywatności. Rewelacje ujawnione przez Edwarda Snowdena⁵ czy skandal z Cambridge Analytica pokazały skalę nadużyć potajemnie zebranych danych internetowych, wykorzystywanych do manipulacji politycznych. Doprowadziło to do międzynarodowej debaty wzywającej do wprowadzenia rygorystycznych przepisów dotyczących przejrzystości regulacji i ochrony informacji, którymi użytkownicy internetu dzielą się on-line.

Wydarzenia te były czymś więcej niż symbolicznym końcem internetowej beztrioski, ujawniły skalę erozji cyfrowej prywatności i makiaweliczne strategie nadużyć, których celem jest zdobycie władzy i kapitalizacja zysków za wszelką cenę. Nie przestaliśmy co prawda korzystać z internetu, ale wzrosła nasza świadomość dotycząca potencjalnych zagrożeń.

Sieć to nadal miejsce wywierania wpływu i dotarcia do konkretnych odbiorców, aby rozbudzić ich ciekawość, zdobyć zaangażowanie, uwagę i długoterminowo – lojalność. Henry Jenkins zdefiniował takie działania jako *ekonomię afektywną*, w ramach której zaciera się granica między przekazem komercyjnym a treścią neutralną, często o charakterze rozrywkowym. Wpływ mediów społecznościowych na biznes jest niezaprzeczalny, dały one początek influencerom, którzy potrafią zgromadzić ogromne rzesze obserwatorów,

⁵ A. Turner, M.W. Zieliński, *Zainteresowanie opinii publicznej tematyką inwigilacji, prywatności i ochrony danych. Google big data w socjologii porównawczej*, „Przegląd Socjologiczny” 2021, t. 70, z. 2, s. 129–156.

co z kolei zrewolucjonizowało strategie marketingowe, umożliwiając firmom bezpośredni kontakt z konsumentami. Stworzyło to nowe możliwości wyrażania siebie, ale wzbudziło także obawy o autentyczność i wpływ na zdrowie psychiczne, szczególnie wśród młodszych użytkowników internetu, na których wpływ mają popularni influencerzy. Ciągłe bycie on-line i wszechobecna kultura porównań mogą się przyczyniać do niepokoju, kompleksów i poczucia niewystarczalności. Ponadto powszechność cyberprzemocy wzbudza poważne obawy, a w rezultacie prowadzi do dyskusji na temat bezpieczeństwa w internecie oraz odpowiedzialności platform za radzenie sobie ze szkodliwymi treściami.

W tej wielowektorowej rzeczywistości przepływów treści między różnymi platformami coraz większa część społeczeństwa porusza się niemal intuicyjnie, zgrabnie nawigując wśród przeróżnych płaszczyzn medialnych, elastycznie wchodząc w rolę to pasywnego odbiorcy, to aktywnego poszukiwacza cyfrowych przygód i informacji. Tak charakteryzuje to zjawisko Zygmunt Bauman:

Nigdy przedtem stare antropologiczne pojęcie „dyfuzji bodźca” (oznaczającego prototypy i inspiracje przenikające lądy i kultury bez pierwotnych wyznawców czy pośredników bądź niezależnie od nich, i poza swoim „naturalnym siedliskiem”), (...) tak dobrze nie wyrażało charakteru współczesnej komunikacji międzykulturowej i zaraźliwego, epidemicznego potencjału innowacji kulturowych. Na planecie poprzecinanej infostradami, [sic!] komunikaty znajdują sobie i wybiorą swoich wdzięcznych słuchaczy, nawet ich nie wyszukując; a raczej same zostaną niezawodnie znalezione i wybrane przez swoich potencjalnych i wdzięcznych słuchaczy, którzy chętnie wezmą na siebie obowiązek szukania („surfowanie po sieci”)⁶.

W tworzenie interaktywnej przestrzeni zmieniającego się informacyjnego spektaklu angażuje się wielu aktorów: regularni bywalcy „świata cyfry”, instytucje, organizacje, firmy i internetowe korporacje, które nieustannie dopracowują monetyzowanie zachowań użytkowników, wykorzystując w tym celu coraz nowocześniejsze technologie. Ponadto, jak pisze socjolog John

⁶ Z. Bauman, *Płynny lęk*, dz. cyt., s. 217–218.

B. Thompson, media internetowe, prezentując wiadomości dwadzieścia cztery godziny na dobę i docierając do szerokiej publiczności, zwiększyły zasięg debaty publicznej⁷. Wchodząc z nim w polemikę, można zadać następujące pytania: Czy komunikacja za pośrednictwem mediów elektronicznych jest współczesną sferą publiczną? Jeśli tak, to na ile jest ona dostępna dla publiczności, a kto jest z niej wykluczony? Spróbuję na te pytania odpowiedzieć.

Wykluczenie informacyjne i cyfrowe

Zwróćmy uwagę na kilka spraw. Korzystanie z internetu różnicują niektóre cechy socjodemograficzne, takie jak: wiek, wykształcenie, a także dostęp do komputera lub telefonu z dostępem do sieci. Ponadto duże znaczenie ma poziom kompetencji cyfrowych, uwzględniających umiejętność wyszukiwania informacji czy adaptacji do stale modyfikowanych zasobów internetowych, takich jak media społecznościowe czy aplikacje⁸.

Z badań panelowych POLPAN wynika, że Polacy przekonują się sukcesywnie do nowych technologii. W roku 2003 ponad 7 na 10 osób nie korzystało z internetu (71,8%), w roku 2018 wciąż ponad jedna czwarta (27,8%). Bez wątplenia prym wiodą najmłodszy – mniej niż 1% respondentów (0,4%) w wieku 18–24 lat nie korzysta z internetu (spadek z 31,9% w roku 2003). Wyniki wśród najstarszych respondentów co prawda się poprawiają, ale bardzo powoli. Aż 96,8% osób w wieku 76+ nie korzystało z internetu w roku 2003. W roku 2018 to wciąż ponad 80% (83,6%), seniorzy są zatem najbardziej wykluczoną cyfrowo grupą Polaków. Znikają różnice między mieszkańcami miast i wsi; w roku 2003 z internetu nie korzystało prawie 9 na 10 mieszkańców wsi (86%), 15 lat później już tylko 3 osoby na 10 (31,4%) – to największy spadek spośród badanych grup, bo aż o 54,6 punktu procentowego. Reasumując: pomimo zmniejszających się różnic mamy jednak do czynienia ze stagnacją wykluczonych cyfrowo – co prawda w perspektywie ostatnich 15 lat zachodzą pewne zmiany, ale bardzo powoli – to nadal

⁷ J. B. Thompson, *Media i nowoczesność. Społeczna teoria mediów*, Astrum, Wrocław 2006.

⁸ Jenkins definiuje te pojęcia jako konwergencję mediów i kulturę uczestnictwa; H. Jenkins, *Kultura konwergencji*, Wydawnictwa Akademickie i Profesjonalne, Warszawa 2006.

najczęściej osoby starsze, z niskim wykształceniem⁹. Warto także zdać sobie sprawę, że dla tych osób kluczowym źródłem informacji są prasa i media analogowe. Tymczasem coraz więcej mediów skupia się na rozwoju subskrypcji cyfrowych. Tylko w roku 2022 kluczowe koncerny prasowe, takie jak: Bauer, Agora, Burda Media Polska i Ringier Axel Springer Polska, zamknęły łącznie ponad trzydzieści tytułów prasowych. Rewolucyjne zmiany zapowiadają się także w radiu; Krajowa Rada Radiofonii i Telewizji jako termin wyłączenia nadawania analogowego przewiduje rok 2030. Przypomnijmy, że w Polsce około 90% Polaków słucha radia na falach FM, nie tylko w domu, lecz także w pracy czy w samochodzie. W momencie gdy radio analogowe przestanie działać, konieczny będzie zakup nowych odbiorników, przystosowanych do odbioru sygnału radiowego nadawanego cyfrowo. Dla wielu osób, które znajdą się w ciężkiej sytuacji ekonomicznej, będzie to bardzo trudne lub niemożliwe, co doprowadzi do jeszcze większej izolacji i zwiększenia wykluczenia informacyjnego. Nie możemy tego lekceważyć.

W starciu z algorytmami sztucznej inteligencji

Od listopada 2022 r. nasze cyfrowe pole semantyczne zostało rozbudowane o nowe słowa. W niedalekiej przyszłości przynajmniej niektóre z nich będą znane większości użytkowników sieci: ChatGPT, stworzony przez firmę OpenAI (prace finansowała firma Microsoft), Bard i Gemini, stworzone przez Google, Grok, opracowany przez xAI – start-up Elona Muska. Czym są? Mówiąc najprościej, są to narzędzia sztucznej inteligencji, które potrafią generować krótsze lub dłuższe odpowiedzi (nawet w formie eseju) na dowolnie zadane pytania. Systemy te nie potrafią jeszcze rozmawiać dokładnie tak jak człowiek, ale prezentowane przez nie odpowiedzi są udzielane natychmiast i mogą sprawiać takie wrażenie. Można o nich myśleć jak o cyfrowych asystentach – w rodzaju Siri czy Alexa – tylko lepiej rozumiejących pytania i potrzeby użytkownika. Interesującym aspektem tej technologii jest jej uniwersalność, ponieważ może mieć zastosowanie w projektach akademickich, jak i twórczych;

⁹ A. Turner, *Polacy w sieci – cyfrowi tubylcy i wykluczeni*, https://ifispan.pl/wp-content/uploads/2022/06/cyklicznik-2022_3-gotowy.pdf (14.10.2023).

chatboty i podobnie działające modele sztucznej inteligencji są w stanie generować biznesplany, tematy prac zaliczeniowych, scenariusze rozmów rekrutacyjnych, wiersze, przemówienia, eseje, pomysły na badania naukowe, diagnozy medyczne, gry o Harrym Potterze. Mogą także wyjaśniać koncepcje naukowe w sposób zrozumiały zarówno dla kilkulatka, jak i eksperta. Innymi słowy, mogą być nieocenionym wsparciem dla tych, którzy go potrzebują.

W świecie rzeczywistym dostęp do ekspertów jest ograniczony z różnych względów, także finansowych. Obecny sposób korzystania i działania internetu dla wielu osób jest zbyt skomplikowany. Aby znaleźć w wyszukiwarce to, czego się szuka, potrzebna jest umiejętność dobrania właściwych słów kluczowych, a następnie selekcji odpowiedniej strony internetowej, zdjęcia lub filmu. Tymczasem internet jest przesycony informacjami do tego stopnia, że nie jesteśmy w stanie przetworzyć nawet części z nich, a umiejętność korzystania z treści cyfrowych w dużej mierze jest zależna od umiejętności ich znalezienia i selekcji. Potrzebny jest ktoś (coś), kto jest w stanie nad tym zapanować, kto dostarczy potrzebnych informacji w przystępny sposób, co wiąże się zarówno z odpowiedzialnością, jak i władzą nad tym, co zostanie wybrane do pokazania, a co do ukrycia. Dlatego nawet jeśli istnieje pilna potrzeba uporządkowanego przeglądu informacji, to należy pamiętać o potencjalnych konsekwencjach. Przy takim tempie rozwoju sztucznej inteligencji nawet nie zauważymy, kiedy stanie się ona wszechobecna, przenikająca każdy aspekt ludzkiego funkcjonowania, niewidoczna. (Nad)używanie narzędzi sztucznej inteligencji dopiero przybiera na sile, niemniej już jesteśmy bogatsi o wiedzę i doświadczenia gromadzone w ciągu ostatnich trzech dekad. Zdajemy sobie sprawę, że nawet najlepsze technologie stworzone w szczytnym celu mogą być wykorzystane (i zapewne będą) do niszczenia, wykorzystywania, żerowania na słabościach, braku wiedzy i umiejętności, do manipulowania czy to dla zysku finansowego, czy władzy. Internet jest naturalną częścią życia, nawet najmłodszy wchodzi w świat pełen algorytmów bazujących na programach sztucznej inteligencji. Aby z nimi współpracować, musimy znać zasady działania, ich cechy charakterystyczne, mocne i słabe strony.

Systemy sztucznej inteligencji oparte na dużych modelach językowych, specjalizujących się w przetwarzaniu i generowaniu języka naturalnego, są potężnymi narzędziami służącymi do wytwarzania i dostarczania informacji, nie są jednak nieomyślne i nigdy nie powinny być jako takie traktowane,

ponieważ pracują na danych, które przygotował człowiek, a zatem będą produkować błędy, a nawet rasistowskie czy dyskryminujące pomówienia. Programy takie jak ChatGPT nie mają świadomości, poczucia winy, solidarności z innymi, duchowości, intuicji, nie biorą pod uwagę szerszego kontekstu, a jednak ludzie mogą je odbierać jako autorytatywne i wiarygodne źródło prawdy, co niesie ze sobą szereg potencjalnych zagrożeń, zwłaszcza jeśli użytkownicy nie są świadomi ich ograniczeń lub jeśli szukają szybkich odpowiedzi, ale nie potrafią samodzielnie dokonać krytycznej oceny lub weryfikacji dostarczonych informacji. Wymieńmy kilka z tych zagrożeń: wersja ChatGPT 3.5 bazuje na danych zebranych do stycznia 2022 r., co oznacza, że w pewnych obszarach program może generować treści nieaktualne; poleganie na niesprawdzonych, niedokładnych, a także potencjalnie stronicznych danych może prowadzić do mylnej interpretacji, do rozpowszechniania fałszywych lub wprowadzających w błąd treści czy też do nadmiernego uproszczenia złożonych kwestii. Co więcej, użytkownicy mogą nieświadomie ujawniać osobiste lub wrażliwe informacje, nie zdając sobie w pełni sprawy z konsekwencji, czyli utraty kontroli nad przekazanymi danymi.

Ludzki osąd, ostrożność i sceptycyzm powinny zawsze odgrywać kluczową rolę w interpretowaniu treści generowanych przez algorytmy. Dlatego potrzebujemy praktycznego doświadczenia, aby zrozumieć, jak działa ten rodzaj sztucznej inteligencji, jakie rodzaje uprzedzeń zawiera i jak może być wykorzystany do złych celów. To dostosowanie nie będzie łatwe. Ze zmianami technologicznymi o charakterze nagłym mamy do czynienia rzadko. Dlatego tak ważne jest nauczanie i rozmawianie o bezpiecznym i odpowiedzialnym korzystaniu z tych narzędzi, także za pomocą mediów (w tym elektronicznych).

Organy rządowe mają do odegrania kluczową rolę, jeśli chodzi o regulacje prawne. Organizacje i indywidualni eksperci są tu kluczowymi interesariuszami i muszą traktować etyczne implikacje swojej pracy znacznie poważniej, niż odbywa się to obecnie. Ponadto twórcy narzędzi takich jak ChatGPT są odpowiedzialni za promowanie przejrzystości oraz wdrażanie zabezpieczeń w celu zminimalizowania potencjalnych negatywnych skutków. Wszystko sprowadza się także do regularnego użytkownika systemu, ludzi, którzy wchodzi w interakcje z chatbotami i innymi modelami sztucznej inteligencji. Ważne jest, aby rozumieli sposób ich działania.

Nie mam złudzeń, że w świecie stale rozwijających się nowych technologii do rywalizacji o ludzką ciekawość i uwagę stajemy w szranki z algorytmami i sztuczną inteligencją. Widzę w nich ogromny potencjał, ale i rodzaj wyzwania, któremu intelektualiści, w tym akademicy, nauczyciele, dziennikarze, publicyści, muszą wyjść naprzeciw, wykorzystując je w celu lepszego, bardziej efektywnego dotarcia do różnych grup odbiorców i jednoczesnego budowania relacji opartej na zaufaniu, zrozumieniu i ludzkiej interakcji. Uważam, że w dużej mierze od tych działań (lub ich braku) zależy, czy elektronicznie zapośredniczona bliskość zastąpi kontakt z drugim człowiekiem.

Przedstawione przeze mnie rozważania zakończę cytatem, który według mnie dobrze oddaje to, co sama chciałabym powiedzieć. Pochodzi on z najnowszej książki Reida Hoffmana, współzałożyciela portalu LinkedIn, członka zarządu firmy Microsoft, kilku organizacji NGO i start-upów zajmujących się nowymi technologiami, można więc z pewnością uznać go za eksperta i autorytet w tej dziedzinie. Hoffman tak podsumowuje naszą przyszłość współdzieloną ze sztuczną inteligencją:

Paradoks ery sztucznej inteligencji polega na tym, że w miarę jak wciąż jeszcze nieidealne duże modele językowe doskonalą się i stawiają przed nami coraz większe wymagania, my sami będziemy zmuszeni oczekiwać od siebie jeszcze więcej. Musimy dążyć do tego, by GPT-4 i jego następcy symbolizowali naszych partnerów, nie zastępców. Nie ustawajmy w wysiłkach, by rozwijać umiejętność oceny, pielegnować wartości, które stanowią o naszym społeczeństwie, nawet wówczas, gdy sztuczna inteligencja zyskuje coraz większe możliwości. W chwilach zwątpienia pamiętajmy, że współpraca z narzędziami AI może przynieść wiele satysfakcji i nadzwyczajne rezultaty. Zatem możemy użyć sztucznej inteligencji, aby pracować mniej lub aby pracować lepiej. Pracując lepiej, damy ludzkości szansę na rozwój. Mając na uwadze, że droga postępu to ścieżka, którą *homo techné* podąża od czasów wczesnych hominidów, jestem optymistą, że i my obierzemy ten kierunek. Czy jesteśmy gotowi wyruszyć w taką podróż?¹⁰

¹⁰ R. Hoffman, *Rozmowa z chatem GPT o przyszłości ludzi i świata*, Prószyński, Warszawa 2023, s. 270.

Bibliografia

- Bauman Z., *Płynny lęk*, Wydawnictwo Literackie, Kraków 2008.
- Hoffman R., *Rozmowa z chatem GPT o przyszłości ludzi i świata*, Prószyński, Warszawa 2023.
- Jenkins H., *Kultura konwergencji*, Wydawnictwa Akademickie i Profesjonalne, Warszawa 2006.
- Skarga B., *Granice historyczności*, PIW, Warszawa 1989.
- Thompson J. B., *Media i nowoczesność. Społeczna teoria mediów*, Astrum, Wrocław 2006.
- Turner A., *Polacy w sieci – cyfrowi tubylcy i wykluczeni*, https://ifispan.pl/wp-content/uploads/2022/06/cyklicznik-2022_3-gotowy.pdf (14.10.2023).
- Turner A., Zieliński M. W., *Zainteresowanie opinii publicznej tematyką inwigilacji, prywatności i ochrony danych. Google big data w socjologii porównawczej*, „Przegląd Socjologiczny” 2021, t. 70, z. 2, s. 129–156.
- Wróbel Sz., *Interdyscyplinarność jako efekt dyscyplinarności*, w: *Głosy w sprawie interdyscyplinarności*, red. J. Kurczewska, M. Lejzerowicz, Wydawnictwo IFiS PAN, Warszawa 2014.

ChatGPT, BARD, ERNIE – are the algorithms of artificial intelligence the authorities of the digital future?

Summary

Cyberspace has imperceptibly become another dimension of social and cultural life, where web users work, spend their leisure time, enjoy cultural goods and seek social contacts. The development that has made it easier and more convenient to use, but it has also created temptations and risks that have triggered discussions about the extent and quality of its influence on our lives and behaviour. In this article, I focus on three aspects: the consequences of the inevitable transformation of the expert culture of the scientific community, the impact of new technologies on the functioning of the virtual space of influence, and the role of chatbots – artificial intelligence algorithms that can completely change the way we interact with technology. I also want to focus on those who remain informationally and digitally excluded.

Keywords: Internet, artificial intelligence, algorithms, chatbots, digital exclusion, information exclusion

Relacja między człowiekiem a maszyną w kontekście rozwoju sztucznej inteligencji i teologii komunikacji

Wstęp

Ostatni sobór powszechny Kościoła katolickiego w swoim dekrecie o środkach społecznego przekazu pt. *Inter mirifica* wypowiada się w następujący sposób:

[...] wśród niezwykłych wynalazków techniki, które zwłaszcza w najnowszych czasach ludzki geniusz z pomocą Bożą rozwinął spośród rzeczy stworzonych, Matka Kościół ze szczególnym zainteresowaniem przyjmuje i popiera te, które w największym stopniu dotyczą ducha ludzkiego i które otworzyły nowe drogi bardzo łatwego przekazywania wszelkich informacji, przemyśleń i zleceń².

Słowa te, jak również cały dokument mogły się wydawać w czasie swojego powstawania, tj. w latach 60. XX wieku, niepotrzebne i zbyteczne. Jaki jest sens tego, aby Kościół jako instytucja religijna miał się wypowiadać o technologii, sposobach komunikacji i przekazywania wiedzy?

¹ Mgr Hubert Siepert, doktorant, Wydział Teologiczny UKSW.

² Sobór Watykański II, *Dekret o środkach społecznego przekazu*, nr 1, w: *Sobór Watykański II. Konstytucje, dekryty, deklaracje*, Pallottinum, Poznań 2002.

Powagę oraz proroczy charakter tego dokumentu i jego przesłania można zrozumieć dopiero w czasach współczesnych. Współczesna kultura w dużym stopniu jest tworzona przez media społecznościowe i świat cyfrowy³. Od ostatnich kilku lat natomiast (a w sposób zdecydowany od listopada 2022 r., kiedy to firma OpenAI wypuściła do powszechnego dostępu ChatGPT 3.5)⁴, w rzeczywistości społecznej pojawiają się coraz bardziej wyspecjalizowane systemy sztucznej inteligencji (*artificial intelligence*; AI), oparte na wielkich systemach językowych (*large language models*; LLMs)⁵. Pozwalają one na bardziej płynną i lepszą komunikację i generują nową jakość w relacji między człowiekiem a maszyną, gdyż coraz lepiej symulują ludzki sposób komunikacji w mowie i piśmie⁶. Owe narzędzia będą coraz lepiej imitować człowieka i ludzkie zachowania, sprawiać wrażenie, że żyją, mają świadomość i swój świat wewnętrzny. Oprócz tego ich struktura fizyczna, na której będą się zasażać, coraz bardziej będzie przypominać ludzką. Wskazuje na to projekt szwajcarskiego start-upu FinalSpark, planującego wprowadzić na rynek nowe bioprocesory, które będą w sobie integrować elementy digitalowe z organicznymi, ludzkimi neuronami⁷. Te i podobne problemy będą pociągać za sobą pytania o to, jak traktować coraz bardziej rozwinięte i zaawansowane systemy AI.

Autor niniejszego artykułu postara się nakreślić pewne wyzwania, które stoją przed człowiekiem w kwestii jego relacji z coraz bardziej wyspecjalizowanymi systemami AI. Na początku zostaną przedstawione antropologiczne podstawy ludzkich tendencji do antropomorfizowania świata i wykrywania

³ Ten wirtualny świat określa się mianem cyfrowego kontynentu; por. M. Flader, *Inkulturacja wiary w „cyfrowy kontynent”*, „Teologia i Człowiek” 2019, t. 47, nr 3, s. 136–139.

⁴ Por. OpenAI Blog, *Introducing ChatGPT*, <https://openai.com/blog/chatgpt> (11.12.2023).

⁵ Jest to jeden z wymiarów szerokiej gamy modeli językowych obok statystycznych modeli językowych, treningowych modeli językowych i neuronalnych modeli językowych; por. M.U. Hadi *et al.*, *A Survey on Large Language Models: Applications, Challenger, Limitations, and Practical Usage*, s. 2, https://www.researchgate.net/publication/372258530_Large_Language_Models_A_Comprehensive_Survey_of_its_Applications_Challenges_Limitations_and_Future_Prospects#fullTextFileContent (11.12.2023).

⁶ Mowa tu o różnego rodzaju botach, chatbotach, voicebotach, które są w stanie imitować człowieka i jego zachowanie; por. A. K. Przegalińska, *Istoty wirtualne. Jak fenomenologia zmieniała sztuczną inteligencję*, Universitas, Kraków 2016, s. 234.

⁷ Por. FinalSpark, *Biocomputing – The next evolutionary leap*, 13.07.2023, <https://finalspark.com/biocomputing-the-next-evolutionary-leap/> (22.11.2023).

w nim intencjonalnie działających agentów. Następnie zostanie przybliżona charakterystyka coraz bardziej zaawansowanych systemów sztucznej inteligencji bazujących na przetwarzaniu naturalnego języka ludzkiego, co powoduje coraz większą ich antropomorfizację przez ludzi. Ostatnim elementem będzie przedstawienie teologii komunikacji w ujęciu Kościoła katolickiego oraz wykazanie w jej świetle, że nie jest uprawnione, aby relacje z systemami sztucznej inteligencji traktować jako komunikację w głębokim, interpersonalnym znaczeniu tego słowa.

Naturalne predyspozycje człowieka do życia społecznego

Sformułowanie, że człowiek jest bytem społecznym⁸, wydaje się tak oczywiste i tak często powtarzane, że staje się swoistym truizmem. Owo określenie człowieka pojawia się w naukach społecznych, refleksji filozoficznej, jak również coraz częściej w naukach biologicznych, powiązanych z biologią ewolucyjną, kognitywistyką i neurobiologią. Swoiste połączenie nauki społecznej z naukami biologicznymi pociąga za sobą zaistnienie neurokognitywnej nauki społecznej, która dąży do odkrycia „kognitywnego wyposażenia niezbędnego dla trwania i rozwoju poszczególnych form życia społecznego ustanowionych przez ewolucję biologiczną”⁹. Człowiek w perspektywie całego gatunku stanowi pochodną procesów ewolucyjnych, które ukierunkowały go na życie społeczne. Z tego też tytułu został on wyposażony w pewne umiejętności społeczne, wynikające z jego bagażu ewolucyjnego.

Jedną z takich umiejętności jest zdolność identyfikacji drugiej osoby na podstawie układu poszczególnych i niezmiennych własności, jak np. twarz, postura, umiejętność mowy¹⁰. Innym mechanizmem, który pojawia się

⁸ Por. P. Przybysz, *Refleksja a stopnie poznania społecznego. W stronę kognitywnej koncepcji umysłu społecznego*, w: *Funkcje umysłu*, red. M. Urbański, P. Przybysz, Zysk i S-ka, Poznań 2009, s. 135.

⁹ Tamże, s. 136.

¹⁰ Por. tamże. s. 137.

u człowieka dość spontanicznie i szybko, jest przypisywanie intencjonalnego działania obiektom różnego rodzaju, zarówno ożywionym, jak i nieożywionym, traktowanym w ten sposób jako wolny, działający w świecie intencjonalny agent¹¹. Takie podejście jest we współczesnej filozofii umysłu określane jako potoczna psychologia pragnień i przekonań, tj. psychologia intencjonalna¹². W świetle interakcji z systemami sztucznej inteligencji kontakt z chatbotem lub voicebotem, który potrafi do nas pisać, rozumieć nasz język bądź mówić do nas naturalnie brzmiącym językiem, może generować na płaszczyźnie psychologicznej upodmiotawianie takiego urządzenia.

Człowiek w swoich relacjach społecznych jest zdolny do tego, co określa się detekcją stanów wewnętrznych innych podmiotów¹³. Ta zdolność pomaga ludziom w rozpoznawaniu intencji innych podmiotów, pozwala współodczuwać, osądzać czyny innych oraz podejmować współpracę. Ludzie są w stanie przez obserwację innych przypisywać im określone stany mentalne i przewidywać ich zachowania¹⁴. Zdolność tę nazwano teorią umysłu (*theory of mind*). Określa się ją jako zdolność „do przypisywania oraz dokonywania atrybucji nieobserwowalnych stanów mentalnych (stanów poznawczych, takich jak przekonania, oraz stanów emocjonalnych, takich jak pragnienia, emocje) w celu przewidywania oraz wyjaśniania zachowań innych osób”¹⁵. W świetle teorii umysłu człowiek jest w stanie zrozumieć i wczuć się w stany mentalne drugiej osoby, określić jej predyspozycje wewnętrzne oraz zrozumieć, że ona również może mieć swoje przekonania, sądy i opinie, które mogą być prawdziwe bądź fałszywe.

Należy zadać w tej materii jedno pytanie: Czy człowiek opisujący drugi byt, np. drugą osobę ludzką, a w przyszłości antropoidalnego robota bazującego na zaawansowanych systemach AI, powinien wiedzieć, że ma kontakt z podmiotem, który myśli i pragnie, czy też wystarczy, aby mu tego typu

¹¹ Por. B. Trybulec, *Podmiot czy agent? Rozumienie podmiotowości w erze artefaktów poznawczych*, „Filozofia i Nauka. Studia filozoficzne i interdyscyplinarne” 2020, t. 8, nr 2, s. 90.

¹² Por. K. Paprzycka, *Fałszywa świadomość psychologii intencjonalnej. O pewnym dogmacie współczesnej analitycznej filozofii umysłu*, „Filozofia Nauki” 2003, t. 41, nr 1, s. 90.

¹³ Por. P. Przybysz, *Refleksja a stopień poznania...*, dz. cyt., s. 142.

¹⁴ Por. A. Pluta, *Mechanizmy poznawcze teorii umysłu*, „Roczniki Psychologiczne” 2012, t. 15, nr 1, s. 8.

¹⁵ Tamże.

myśli i pragnienia przypisał? Jeśli przyjmie się tę drugą opcję, to wówczas można powiedzieć, że „system nie jest traktowany jako podmiot postaw propozycyjalnych dlatego, że faktycznie je posiada, przeciwnie, staje się on ich podmiotem dlatego, że tak jest traktowany”¹⁶. Stanowisko takie może być pojmowane jako dość instrumentalistyczne podejście do zagadnienia podmiotowości. Nie chodzi w nim o to, aby orzec prawdę o wewnętrznych stanach mentalnych drugiego bytu, z którym wchodzimy w relacje, ale by traktować go jako byt mający intencje i podejmujący decyzje, przez co będzie możliwe przewidywanie jego działania¹⁷.

W takim kluczu nie ma znaczenia, czy bot, voicebot, android czy inny system sztucznej inteligencji coś czuje bądź myśli. Ważne jest, że ludzie przypisują mu tego typu wewnętrzne stany mentalne. Przez taki akt uznania, w ludzkiej subiektywności, takie formy nowych technologii mogą z czasem stawać się towarzyszami ludzkiego życia, partnerami do zwierzeń albo pełnić inne socjalne bądź quasi-socjalne funkcje. Poprzez analogię – tak jak współcześnie ludzie dokonują antropomorfizacji swoich zwierząt, tak też i w przyszłości mogą jej dokonywać względem systemów sztucznej inteligencji. Co więcej, pojawiają się także głosy, że:

[...] zjawisko antropomorfizacji zwierząt jest bardzo powszechne, konieczne, a wręcz – z punktu widzenia człowieka – niezbędne, by móc obcować ze zwierzętami w codziennym życiu. Antropomorfizowanym zwierzętom najczęściej nadawane są cechy o znaczeniu psychologicznym oraz społecznym. Umysł ludzki intuicyjnie obdarza większym znaczeniem oraz etyczną wartością te istoty, które „w jego opinii” mają „życie wewnętrzne”¹⁸.

Tak jak współcześnie ludzie uczłowiczają swoje zwierzęta domowe, tak przez analogię będzie może w przyszłości postępowało „uczłowiczanie” systemów sztucznej inteligencji.

¹⁶ B. Trybulec, *Podmiot czy agent?...*, art. cyt., s. 95.

¹⁷ Por. Tamże.

¹⁸ K. Adamczyk, D. Zwierkowska, *Antropomorfizacja zwierząt – dylemat współczesnych relacji człowiek–zwierzę*, „Przegląd Hodowlany”, 4 (2021), s. 25.

Komunikacja z systemami sztucznej inteligencji

We współczesnym świecie ludzkość coraz częściej wchodzi w szeroko pojętą komunikację z własnymi twórcami, tj. z nowoczesnymi urządzeniami, które wykonują za nas liczne czynności, jak chociażby z komputerami¹⁹. Należy także zwrócić uwagę, że sposób komunikowania i przekazywania informacji do urządzeń informatycznych ulega daleko idącym zmianom. Widać to na przykładzie systemów sztucznej inteligencji. Tworzą one właśnie systemy, tj. zbiory różnych elementów powiązanych wzajemnymi relacjami, które razem biorą udział w realizowaniu pewnych procesów zmierzających do uzyskania konkretnych efektów, jak wytworzenie określonych zjawisk²⁰.

Obecnie, głównie dzięki *natural language processing* (NLP), które należy rozumieć jako technologię umożliwiającą komputerom zrozumienie i przetwarzanie tekstu i mowy w języku naturalnym, dzięki czemu „możemy korzystać z różnych aplikacji, takich jak tłumaczenie maszynowe, rozpoznawanie mowy, analiza sentymentu, automatyczne odpowiedzi na e-maile i wiele innych”²¹, interakcja ze sztuczną inteligencją staje się coraz bardziej powszechna i dostępna w użytkowaniu. Technologia ta, w połączeniu z systemami sztucznej inteligencji, a szczególnie z generatywną sztuczną inteligencją, pozwala na coraz lepsze symulowanie postaw i zachowań ludzkich i podmiotowych w interakcjach. Wynika to z tego, że system AI, np. ChatGPT, pojmuje i przetwarza ludzki język, a także, jako generatywna AI: „wykorzystuje algorytmy uczenia maszynowego do generowania nowych, oryginalnych i podobnych do ludzkich treści. Jest to możliwe dzięki dwóm sieciom neuronowym, które tworzą GAN (generatywne sieci przeciwstawne)”²².

¹⁹ Komputer jest maszyną symulującą, która przez matematyczną symulację pełni funkcje urządzenia, które akurat symuluje; por. H. Rfeingold, *Narzędzia ułatwiające myślenie – historia i przyszłość metod poszerzania możliwości umysłu*, tłum. J. B. Szporko, Warszawa 2003, s. 16.

²⁰ Por. B. Trybulec, *Podmiot czy agent?...*, art. cyt., s. 91.

²¹ The Force Code, *Jak NLP i sztuczna inteligencja zmieniają komunikację?*, <https://theforcecode.com/help/pl/jak-nlp-i-sztuczna-inteligencja-zmieniaja-komunikacje/> (27.11.2023).

²² PART. Grupa PFR, System Rad ds. Kompetencji, *Rynek pracy, edukacja, kompetencje. Wykorzystanie sztucznej inteligencji w edukacji*, wyd. specjalne, <https://www.parp.gov.pl/storage/publications/pdf/Wykorzystanie-sztucznej-inteligencji-w-edukacji.pdf>, s. 18 (27.11.2023).

Dziedzina NLP w sposób zdecydowany zmienia ludzką formę komunikacji z wytworami świata informatycznego i cyfrowego. Dzięki tej technologii możliwa jest komunikacja z maszynami, które stanowią coraz bardziej zaawansowane systemy. Co więcej, są one w stanie, dzięki rozpoznawaniu ludzkiego języka naturalnego, odczytywać intencje i udzielać odpowiedzi na nasze pytania, przez co komunikacja z systemami AI staje się o wiele bardziej naturalna i intuicyjna²³.

Kolejnym problemem jest to, że owe technologie cały czas się rozwijają i stają się coraz bardziej zaawansowane. Wspomniany wcześniej ChatGPT, dzieło firmy OpenAI, zaczyna mieć prawdziwego konkurenta, którym staje się AI od Google, tj. Gemini. Jak podają jej twórcy, jest to pierwszy system AI, który przewyższył ludzkich ekspertów w MMLU (*massive multitask language understanding*)²⁴. W ramach tej rywalizacji Gemini (w wersji ultra, która ma być dostępna w 2024 r.) osiągnął poziom 90%, ludzcy eksperci – 89,9%, poprzednie modele natomiast, takie jak ChatGPT 4, osiągnął poziom 86,4%²⁵. W dniu 6 grudnia 2023 r. został też na portalu YouTube zaprezentowany film, który pokazuje zdolności multimodalne Gemini AI Ultra i ich wykorzystanie w integracji z człowiekiem²⁶. Multimodalność można rozumieć jako zdolność nie tylko do łączenia różnych dziedzin, wyciągania wniosków i sprawnego poruszania się po zagadnieniach i materiałach źródłowych, lecz także do radzenia sobie z jednoczesnym interpretowaniem zdjęć, tekstu, dźwięku czy kodu komputerowego²⁷. Gemini AI nie jest zatem modelem językowym generatywnej sztucznej inteligencji, który bazowo miał być chatbotem. Od samego początku był trenowany na różnego rodzaju danych treningowych, takich jak tekst, obraz, dźwięk, bazy danych.

²³ Por. The Force Code, *Jak NLP i sztuczna inteligencja...*, art. cyt.

²⁴ Por. Google DeepMind, *Welcome to the Gemini era*, <https://deepmind.google/technologies/gemini/#introduction> (8.12.2023).

²⁵ Por. tamże.

²⁶ Por. YouTube, *Hands-on with Gemini: Interacting with multimodal AI*, <https://www.youtube.com/watch?v=UIZAiXYceBI>, (8.12.2023).

²⁷ Por. D. Szafraniak, *Czym jest Google Gemini? Rewolucja od Google dla każdego*, <https://geekweek.interia.pl/technologie/news-czym-jest-google-gemini-rewolucja-od-google-dla-kazdego,nId,7195145> (8.12.2023).

Biorąc pod uwagę owe przemiany oraz swoistą naturalną skłonność ludzkości do ucłowieczania tego, co z pozoru przypomina człowieka i komunikuje się jak człowiek, istnieje pewne ryzyko tego, że coraz bardziej zaawansowane systemy AI będą rzeczywiście wprowadzać ludzi w błąd co do swojej tożsamości. Jeśli takie systemy ograniczałyby się tylko do tego, aby wyręczać ludzi w ciężkich, niebezpiecznych bądź nużących pracach, to takie rozwiązania mogą rzeczywiście stać się błogosławieństwem dla rodzaju ludzkiego.

Jednakże większym problemem będzie to, gdy systemy sztucznej inteligencji zaczną wchodzić w relacje zachodzące między ludźmi bądź też je zastępować, nie tylko na płaszczyźnie wiedzy i kompetencji, ale również w wymiarze empatyczności. Dobrym przykładem takiej sytuacji jest badanie przeprowadzone przez naukowców z Uniwersytetu Kalifornijskiego w San Diego, którzy przeanalizowali 195 pytań zadanych lekarzom w USA na forum AskDocs²⁸. Te same pytania zostały również zadane przez pacjentów chatowi GPT. Pacjenci ci nie wiedzieli, czy komunikują się z chatbotem, czy z prawdziwym lekarzem. Z badania tego wynika, że odpowiedzi chatu GPT były preferowane częściej niż odpowiedzi lekarzy osobowych. Reakcje systemu były traktowane również jako bardziej empatyczne²⁹. Nie znaczy to, że sam ChatGPT jest rzeczywistością empatyczną, podmiotową i moralną, wskazuje jednak, że może być przez swoich odbiorców w taki sposób traktowany. Dodatkowo trzeba założyć, że jest on dość wczesnym przedstawicielem technologii generatywnej AI.

Z perspektywy eklezjalnej problem ten zauważył w 2023 r. kardynał arcybiskup Utrechtu oraz prymas Niderlandów Willem Jacobus Eijk. Podkreślił on konieczność przemyślenia interakcji między człowiekiem a systemami sztucznej inteligencji oraz robotami. Zwrócił uwagę w sposób szczególny na sferę opieki nad osobami starszymi i chorymi, gdzie to roboty mogą być wykorzystywane np. do karmienia takich osób, co poskutkuje dehumanizacją tejże sfery i może pozbawić pensjonariuszy ośrodków pewnego

²⁸ Por. Centrum Nauki Eksperyment, *Czy sztuczna inteligencja jest bardziej empatyczna od lekarza?*, <https://experyment.gdynia.pl/dzialania-on-line/artykuly-popularnonaukowe/tajemnice-umyslu/czy-sztuczna-inteligencja-jest-bardziej-empatyczna-od-lekarza/> (8.12.2023).

²⁹ Por. J. W. Ayers *et al.*, *Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum*, „JAMA Internal Medicine”, s. 6, https://kstp.com/wp-content/uploads/2023/05/jamainternal_ayers_2023_oi_230030_1681999216.70842.pdf (8.12.2023).

kontaktu międzyludzkiego³⁰. Należy także, w ujęciu kardynała, wprowadzać maszyny do życia społecznego w sposób rozsądny i nawiązywać z nimi interakcje tak, aby nie prowadziło to do pogłębiania poczucia samotności, które przejawia się we współczesnych społeczeństwach rozwiniętych³¹.

Do refleksji nad roztropnym wprowadzeniem systemów AI w życie społeczne wzywa również papież Franciszek. W swoim Orędziu na 58. Światowy Dzień Środków Społecznego Przekazu pisze następujące słowa:

W tej epoce, której grozi, że będzie bogata w technologię a [sic!] uboga w człowieczeństwo, nasza refleksja może wychodzić jedynie od ludzkiego serca. Tylko mając spojrzenie duchowe, tylko odzyskując mądrość serca, możemy odczytać i zinterpretować nowość naszych czasów i na nowo odkryć drogę do komunikacji w pełni ludzkiej³².

Pojawienie się myślących i posługujących się naturalnym językiem maszyn staje się w ujęciu papieża Franciszka wezwaniem do tego, aby na nowo odkryć głębię i bogactwo człowieczeństwa, zarówno w swoim wnętrzu, jak i w wymiarze szeroko pojętego życia społecznego.

Refleksja natury teologicznej nad komunikacją człowiek–maszyna

Na relację, jaka się pojawia między człowiekiem a stworzonymi przez niego coraz bardziej autonomicznymi artefaktami kultury technologicznej, można spojrzeć z perspektywy teologii komunikacji. Jest to nurt teologii bardziej powszechny na Zachodzie i dotyczy „komunikacji medialnej, czy poza

³⁰ Por. Vatican New, *Kard. Eijk: potrzeba refleksji Kościoła wobec sztucznej inteligencji*, <https://www.vaticannews.va/pl/kosciol/news/2023-05/kard-eijk-potrzeba-refleksji-kosciola-wobec-sztucznej-intelige.html> (9.12.2023).

³¹ Por. tamże.

³² Franciszek, Orędzie na 58. Światowy Dzień Środków Społecznego Przekazu, *Sztuczna inteligencja i mądrość serca: dla komunikacji w pełni ludzkiej*, <https://www.vatican.va/content/francesco/pl/messages/communications/documents/20240124-messaggio-comunicazioni-sociali.html> (24.03.2024).

medialnej [*sic!*], a więc bezpośredniej komunikacji interpersonalnej, indywidualnej oraz zbiorowej, a także indywidualnej i zbiorowej komunikacji dokonującej się za pośrednictwem rozmaitych urządzeń technicznych³³. Przedmiotem teologii komunikacji jest teologiczna interpretacja zjawiska i całego procesu komunikowania się³⁴.

Przyglądając się powyższej definicji, można byłoby poszerzyć zakres zainteresowania teologii komunikacji. Jej przedmiotem powinna się stać również systematycznie i teologicznie zinterpretowana relacja zachodząca między ludźmi a ich tworamii cyfrowymi, które wykazują się coraz większym zrozumieniem człowieka i możliwością wchodzenia z nim w interakcje. W sposób szczególny powinna być podjęta refleksja nad złożonymi systemami sztucznej inteligencji, zarówno w formie cyfrowych botów, jak i robotów antropoidalnych.

Samo pojęcie komunikacji etymologicznie pochodzi od łacińskiego terminu *communicatio*, które można rozumieć jako ‘udzielanie się’, ‘zawiadomienie’, ‘bycie z kimś w relacji’³⁵. Czasownik *communicare* natomiast oznacza tyle co ‘zawiadomić’, ‘być z kimś w relacji’, ‘uczestniczyć w czymś’³⁶. Przyglądając się etymologii terminu „komunikacja”, można stwierdzić, że ma ono dwa wymiary. Pierwszym jest przekazywanie informacji bądź komend lub poleceń. Ten wymiar ma charakter bardziej przedmiotowy. Drugi natomiast to wymiar wspólnotowy, relacyjny, który polega na przekazywaniu w akcie komunikacji siebie i odbieraniu drugiego. Jest to ujęcie bardziej podmiotowe procesu komunikowania się.

Teologia katolicka rozwija pojęcie komunikacji i nadaje mu głębszy sens. Komunikacja znajduje się w samym jądrze chrześcijaństwa, gdyż wypływa z koncepcji Objawienia, czyli ukazania się Boga człowiekowi. Jak zauważa kard. Gerhard Müller w swojej pozycji pt. *Dogmatyka katolicka*:

³³ M. Butkiewicz, G. Łęcicki, *Ewolucja i specyfika teologii mediów w systemie nauk teologicznych i nauk o mediach*, w: *Teologia środków społecznego przekazu w naukach o mediach*, red. J. Olędzki, Warszawa 2018, s. 100.

³⁴ Por. P. Gabara, *Recepcja zjawiska komunikacyjnego w teologii i homiletyce*, „Łódzkie Studia Teologiczne” 2014, t. 23, nr 3, s. 79.

³⁵ Por. J. A. Sobkowiak, *Komunikacja przejawem podmiotowości osoby*, w: *Teologia środków społecznego przekazu...*, dz. cyt., s. 57.

³⁶ Por. tamże.

Objawienia nie interpretuje się ani jako systemu prawdziwych twierdzeń, ani stanów uczucia religijnego, lecz jako wydarzenie relacji między dwoma podmiotami. Wiara opiera się na pierwotnym doświadczeniu rzeczywistości i na osobistym spotkaniu z Bogiem w jego słowie i działaniu [...] Bosko-ludzkie spotkanie dokonuje się na sposób dialogowy i historyczny. Właśnie w ten sposób zostaje ono objęte specyficzną dla objawienia kategorią „spotkania międzyosobowego”³⁷.

Sam Bóg objawia się człowiekowi i staje się dla niego najdoskonalszym wzorcem komunikowania – komunii przez swoją wewnętrzną trynitarną tajemnicę³⁸.

Z kolei dokument pt. *Communio et progressio* określa, że: „Komunikowanie to coś więcej niż samo tylko przekazywanie idei czy wyrażanie uczuć, to przede wszystkim dawanie siebie w miłości”³⁹. I dalej ten sam dokument kontynuuje refleksję nad fenomenem komunikacji w następujący sposób: „ustanawiając Eucharystię, Chrystus dał nam najwspanialszy wzór «komunikowania», jaki jest dostępny ludziom na ziemi, mianowicie, zjednoczenie, komunikowanie między Bogiem i człowiekiem, a przez to samo wzór najściślejszego i najdoskonalszego zjednoczenia między ludźmi”⁴⁰. Dlatego też teologiczny namysł nad fenomenem komunikacji musi mieć charakter chrystocentryczny. Chrystus rozumiany jest jako *Perfectus Communicator*, tj. Doskonały Mistrz Komunikacji, który stanowi wzór wszelkiego komunikowania z Bogiem i między ludźmi⁴¹.

W ujęciu teologii katolickiej akt komunikacji jest możliwy tylko między dwiema osobami. Komunikacja w najgłębszym wymiarze oprócz charakteru treściowego, który zawiera jakiś przekaz, informację bądź komendę,

³⁷ G. Müller, *Dogmatyka katolicka*, tłum. W. Szymona, Kraków 2015, s. 83–84.

³⁸ Por. J. Szymik, *Teologiczne podstawy rozumienia komunikacji*, s. 37–38, https://dlibra.kul.pl/Content/26845/5_teologiczne_podstawy.pdf (9.12.2023).

³⁹ Papiaska Rada ds. Środków Społecznego przekazu, *Communio et progressio*, nr 11, <https://www.katolickie.media.pl/index.php/polecane/kosciol-o-mediach/414-communio-et-progressio--instrukcja-duszpasterska-o-srodkach-spolecznego-przekazu> (9.12.2023).

⁴⁰ Tamże.

⁴¹ Por. M. Butkiewicz, G. Łęcicki, *Ewolucja i specyfika teologii mediów...*, art. cyt., s. 105; por. Jan Paweł II, List apostolski *Szybki rozwój. O szybkim rozwoju środków społecznego przekazu*, nr 4–5, https://opoka.org.pl/biblioteka/W/WP/jan_pawel_ii/listy/szybkir_24022005 (11.12.2023).

zawiera w sobie także wymiar relacyjny, w ramach którego osoby udzielają się sobie nawzajem⁴². Z tego wynika, że aby prawdziwie zachodził akt komunikacyjny, muszą się spotkać dwa podmioty, które działają intencjonalnie i które można obarczyć odpowiedzialnością moralną⁴³. Komunikowanie tworzy zarówno poznanie, jak i miłość, a tym, co ostatecznie się komunikuje, nie jest treściowym i przedmiotowym komunikatem, ale własną osobą nadawcy⁴⁴.

Z tak postrzeganej wizji komunikacji, zarezerwowanej tylko dla podmiotów moralnych, które charakteryzuje działanie intencjonalne i odpowiedzialność moralna, systemy sztucznej inteligencji zostają wykluczone. Nie przypisuje się im (przynajmniej na chwilę obecną) stanów samoświadomości, takiej jaką mają ludzie⁴⁵. Co więcej, w ramach komunikacji podmiot zawsze udziela siebie jako niepowtarzalny i niezastępowalny. Człowieka, z jego historią, doświadczeniami i przeżyciami, nie da się skopiować ani zgrać na inny nośnik, jak można to uczynić z programami komputerowymi⁴⁶.

Faktem jest, że systemy AI mają coraz doskonalsze możliwości imitowania człowieka i funkcjonowania w ludzkim społeczeństwie. Już w 2018 r. sztuczna inteligencja o nazwie Android Duplex odebrała od człowieka zadanie, aby umówić go na wizytę u fryzjera. System sam wybrał salon oraz zadzwonił do fryzjerki i posługując się kobiecym głosem, umówił się na wizytę. Fryzjerka nie zorientowała się, że rozmówca nie jest człowiekiem⁴⁷. Innym przykładem jest Aitana López, 25-letnia hiszpańska modelka, która zarabia do 10 000 euro miesięcznie, ma 121 tys. obserwujących na Instagramie, została twarzą firmy Bing produkującej suplementy sportowe i publikuje swoje zdjęcia na platformie Fanvue, a przy tym jest botem AI⁴⁸. Wciąż nie zmienia to jednak faktu, że owe boty dostają od innych ludzi zadanie,

⁴² Por. J. A. Sobkowiak, *Komunikacja przejawem...*, art. cyt., s. 61.

⁴³ Por. Tamże, s. 62.

⁴⁴ Por. J. Szymik, *Teologiczne podstawy...*, art. cyt., 38.

⁴⁵ Por. R. Wawer, *Sztuczna inteligencja – produkt czy podmiot?*, „Studia Teologii Dogmatycznej” 2018, t. 4, s. 124.

⁴⁶ Por. Tamże.

⁴⁷ Por. Tamże, s. 125.

⁴⁸ Por. AIOAI, *Aitana López: Pierwsza hiszpańska wirtualna modelka AI*, <https://aioai.pl/aitana-lopez-pierwsza-hiszpańska-wirtualna-modelka-ai/> (9.12.2023).

aby podszyc się pod konkretne osoby. Odgrywają niejako swoje „role”, symulują postawy, gesty, emocje – jednakże nie są osobami.

Można byłoby potraktować tego typu systemy jako swoiste filozoficzne zombie. Pojęcie zombie istnieje w filozofii umysłu i stosuje się je na określenie systemów, które pod względem fizycznym i funkcjonalnym ciężko odróżnić od ludzi, jednak w aspekcie życia psychicznego nie można tu mówić o przeżyciach świadomych.

Mogłyby one prowadzić konwersacje, wykonywać czynności dnia codziennego, uprawiać naukę, seks, tworzyć dzieła sztuki, chociaż nie dysponowałyby żadną formą świadomego doświadczenia. Ich aktywność ograniczałaby się do rządzonych algorytmami automatyzmów. Zewnętrzny obserwator nie potrafiłby ich odróżnić od istot świadomych⁴⁹.

Jednakże w interakcji z takim systemem w sensie dosłownym po „drugiej stronie” nikogo nie ma. Jest tylko ciąg logicznych algorytmów, które uaktualniają się do tego, aby podjąć pewne z góry określone zadanie, wykonać polecenie bądź realizować założone przeznaczenie. Nie będzie z nim autentycznej komunikacji.

Taki system nie będzie doświadczał interakcji z ludźmi. Nie będzie w sposób realny z nimi współcierpiał, cieszył się bądź przeżywał choroby, starość, smutki, bóle czy radości. Może co najwyżej stać się pewną namiastką relacji międzyludzkich, nigdy jednak nie powinien się okazać rzeczywistością zastępującą relacje z prawdziwym drugim człowiekiem.

Jak zauważa bardzo trafnie papież Franciszek: „symulacja, będąca podstawą tych programów, może być przydatna w niektórych konkretnych dziedzinach, ale staje się przewrotna, gdy wypacza relacje z innymi i z rzeczywistością”⁵⁰. Nie możemy zapominać o tym, że nawet najbardziej zaawansowane twory AI są dziełem ludzkich rąk. Są w ostatecznym rozrachunku ludzkimi narzędziami, a nie osobami, przyjaciółmi bądź substytutem jakichkolwiek innych relacji. Tę kwestię papież Franciszek również mocno podkreśla.

⁴⁹ R. Poczobut, *Systemy zombie a świadomość*, „Filozofuj” 2021, nr 1(37), s. 22.

⁵⁰ Franciszek, Orędzie na 57 Światowy Dzień Środków Społecznego Przekazu, *Sztuczna inteligencja i mądrość serca...*, dz. cyt.

Przykłady wskazane powyżej przemawiają za tym, że zaawansowane systemy AI mogą być wykorzystywane w różnych celach, także z zamiarem tworzenia dezinformacji i przekłamań. Dlatego też w nauczaniu papieskim pojawiają się apele o podjęcie kwestii etycznych i prawnych w temacie regulacji nowych technologii: „konieczne jest działanie prewencyjne, proponujące modele regulacji etycznych w celu powstrzymania szkodliwych i dyskryminujących, niesprawiedliwych społecznie implikacji systemów sztucznej inteligencji”⁵¹. Jak każda sfera ludzkiej działalności, także i przestrzeń nowych technologii powinna stać się przedmiotem dyskusji i wypracowania wspólnych rozwiązań o charakterze prawnym i etycznym.

Zakończenie

Człowiek stworzony na obraz i podobieństwo Boga (Rdz 1, 26n) współcześnie co chwila stwarza narzędzia, które coraz lepiej odbijają jego samego. Stanowi to z jednej strony wielki tryumf ludzkiego ducha i intelektu, a z drugiej stawia przed człowiekiem nowe wyzwania i pytania. Jaką rolę w społeczeństwie przyszłości mają odgrywać nowe technologie związane z systemami sztucznej inteligencji, które znakomicie imitują ludzką osobowość i podmiotowość? Autor niniejszego tekstu starał się przedstawić pewne spojrzenie na ową kwestię, która prawdopodobnie z czasem będzie się nasilać.

Na samym początku autor rozdziału nakreślił pewne naturalne predyspozycje, które tkwią w człowieku i są częścią jego ewolucyjnej historii, z jednej strony ułatwiają mu życie społeczne, a z drugiej sprawiając, że człowiek dość łatwo może antropomorfizować i upodmiotawiać byty, które podmiotami nie są. Następnie zostały w artykule przedstawione współczesne trendy rozwoju systemów AI, skupiając się na wielkich modelach językowych i coraz bardziej intuicyjnej relacji, opartej na języku naturalnym, jaka zachodzi między człowiekiem a maszyną. Na końcu została ukazana interpretacja teologiczna tychże zjawisk w kluczu teologii komunikacji. Autor podkreślił, że komunikowanie, rozumiane jako wchodzenie w relacje i udzielanie się

⁵¹ Tamże.

sobie nawzajem, może zachodzić tylko między osobami, które są moralnymi, samoświadomymi podmiotami działania. Z tego typu komunikacji wyłączone są systemy AI, które nie stanowią samoświadomych podmiotów, z tego też tytułu relacje z nimi powinny mieć charakter wspomagający, nie powinny zaś stanowić substytutu relacji międzyludzkich. Przytoczone zostały także na końcu słowa papieża Franciszka z jego Orędzia na 57. Światowy Dzień Środków Społecznego Przekazu, wzywające do głębokiego przemyślenia tych nowych technologii na płaszczyźnie etyki i prawodawstwa.

Bibliografia

- Adamczyk K., Zwierkowska D., *Antropomorfizacja zwierząt – dylemat współczesnych relacji człowiek – zwierzę*, „Przegląd Hodowlany” 2021, nr 4, s. 25–28.
- AIoAI, Aitana López: *Pierwsza hiszpańska wirtualna modelka AI*, <https://aioai.pl/aitana-lopez-pierwsza-hiszpańska-wirtualna-modelka-ai/> (9.12.2023).
- Ayers J. W., Poliak A., Dredze M., Leas E. C., Zhu Z., Kelley J. B., Faix D. J., Goodman A. M., Longhurst Ch. A., Hogarth M., Smith D. M., *Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum*, „JAMA Internal Medicine”, https://kstp.com/wp-content/uploads/2023/05/jamainternal_ayers_2023_oi_230030_1681999216.70842.pdf (8.12.2023).
- Butkiewicz M., Łęcicki G., *Ewolucja i specyfika teologii mediów w systemie nauk teologicznych i nauk o mediach*, w: *Teologia środków społecznego przekazu w naukach o mediach*, red. J. Olędzki, Wydawnictwo UKSW, Warszawa 2018, s. 97–119.
- Centrum Nauki Eksperyment, *Czy sztuczna inteligencja jest bardziej empatyczna od lekarza?*, <https://experyment.gdynia.pl/dzialania-on-line/artykuly-popularnonaukowe/tajemnice-umyslu/czy-sztuczna-inteligencja-jest-bardziej-empatyczna-od-lekarza/> (8.12.2023).
- FinalSpark, *Biocomputing – The next evolutionary leap*, 13.07.2023, <https://finalspark.com/biocomputing-the-next-evolutionary-leap/> (8.12.2023).
- Flader M., *Inkulturowanie wiary w „cyfrowy kontywent”*, „Teologia i Człowiek” 2019, t. 47, nr 3, s. 129–144.
- Franciszek, Orędzie na 57 Światowy Dzień Środków Społecznego Przekazu, *Sztuczna inteligencja i mądrość serca: dla komunikacji w pełni ludzkiej*, <https://www.vatican.va/content/francesco/pl/messages/communications/documents/20240124-messaggio-comunicazioni-sociali.html> (24.03.2024).
- Gabara P., *Recepcja zjawiska komunikacyjnego w teologii i homiletyce*, „Łódzkie Studia Teologiczne” 2014, t. 23, nr 3, s. 77–92.

- Google DeepMind, *Welcome to the Gemini era*, <https://deepmind.google/technologies/gemini/#introduction> (8.12.2023).
- Hadi M. U., Al-Tashi Q., Qureshi R., Shah A., Muneer A., Irfan M., Shaikh M. B., Akhtar N., Wu J., Mirjalili S., *A Survey on Large Language Models: Applications, Challenger, Limitations, and Practical Usage*, https://www.researchgate.net/publication/372258530_Large_Language_Models_A_Comprehensive_Survey_of_its_Applications_Challenges_Limitations_and_Future_Prospects#fullTextFileContent (11.12.2023).
- Jan Paweł II, List apostolski *Szybki rozwój. O szybkim rozwoju środków społecznego przekazu*, https://opoka.org.pl/biblioteka/W/WP/jan_pawel_ii/listy/szybkir_24022005 (11.12.2023).
- Müller G., *Dogmatyka katolicka*, tłum. W. Szymona, Wydawnictwo WAM, Kraków 2015.
- OpenAI Blog, *Introducing ChatGPT*, <https://openai.com/blog/chatgpt> (25.11.2023).
- Papieska Rada ds. Środków Społecznego przekazu, *Communio et progressio*, <https://www.katolickie.media.pl/index.php/polecane/kosciol-o-mediach/414-communio-et-progressio---instrukcja-duszpasterska-o-srodkach-spolecznego-przekazu> (9.12.2023).
- Paprzycka K., *Falszywa świadomość psychologii intencjonalnej. O pewnym dogmacie współczesnej analitycznej filozofii umysłu*, „Filozofia Nauki” 2003, t. 41, nr 1, s. 89–116.
- PART. Grupa PFR, System Rad ds. Kompetencji, *Rynek pracy, edukacja, kompetencje. Wykorzystanie sztucznej inteligencji w edukacji*, wyd. specjalne, <https://www.parp.gov.pl/storage/publications/pdf/Wykorzystanie-sztucznej-inteligencji-w-edukacji.pdf> (27.11.2023).
- Pluta A., *Mechanizmy poznawcze teorii umysłu*, „Roczniki Psychologiczne” 2012, t. 15, nr 1, s. 7–30.
- Poczobut R., *Systemy zombie a świadomość*, „Filozofuj” 2021, nr 1(37), s. 22–24.
- Przegalińska A. K., *Istoty wirtualne. Jak fenomenologia zmieniała sztuczna inteligencję*, Universitas, Kraków 2016.
- Przybysz P., *Refleksja a stopnie poznania społecznego. W stronę kognitywnej koncepcji umysłu społecznego*, w: *Funkcje umysłu*, red. M. Urbański, P. Przybysz, Zysk i S-ka, Poznań 2009, s. 135–153.
- Rfeingold H., *Narzędzia ułatwiające myślenie – historia i przyszłość metod poszerzania możliwości umysłu*, tłum. J. B. Sporko, Wydawnictwo Naukowo Techniczne, Warszawa 2003.
- Sobkowiak J. A., *Komunikacja przejawem podmiotowości osoby*, w: *Teologia środków społecznego przekazu w naukach o mediach*, red. J. Ołędzki, Warszawa 2018, s. 55–80.
- Sobór Watykański II, *Dekret o środkach społecznego przekazu*, w: *Sobór Watykański II. Konstytucje, dekrety, deklaracje*, Pallottinum, Poznań 2002, s. 87–95.
- Szafrański D., *Czym jest Google Gemini? Rewolucja od Google dla każdego*, <https://geekweek.interia.pl/technologie/news-czym-jest-google-gemini-rewolucja-od-google-dla-kazdego,nId,7195145> (8.12.2023).
- Szymik J., *Teologiczne podstawy rozumienia komunikacji*, https://dlibra.kul.pl/Content/26845/5_teologiczne_podstawy.pdf (9.12.2023).

- The Force Code, *Jak NLP i sztuczna inteligencja zmieniają komunikację?*, <https://the-forcecode.com/help/pl/jak-nlp-i-sztuczna-inteligencja-zmieniaja-komunikacje/> (27.11.2023).
- Trybulec B., *Podmiot czy agent? Rozumienie podmiotowości w erze artefaktów poznawczych*, „Filozofia i Nauka. Studia filozoficzne i interdyscyplinarne” 2020, t. 8, nr 2, s. 89–113.
- Vatican New, *Kard. Eijk: potrzeba refleksji Kościoła wobec sztucznej inteligencji*, <https://www.vaticannews.va/pl/kosciol/news/2023-05/kard-eijk-potrzeba-refleksji-kosciola-wobec-sztucznej-intelige.html> (9.12.2023).
- Wawer W., *Sztuczna inteligencja – produkt czy podmiot?*, „Studia Teologii Dogmatycznej” 2018, t. 4, s. 115–126.
- YouTube, *Hands-on with Gemini: Interacting with multimodal AI*, <https://www.youtube.com/watch?v=UIZAIXYceBI> (8.12.2023).

The relationship between man and machine in the context of the development of artificial intelligence and the theology of communication

Summary

The last 250 years of human history have been a period of continuous technological development, which entails transformations at both the social and individual levels. Currently, another civilizational transformation is emerging, linked to the development of new teleinformatics technologies and artificial intelligence systems. This issue should also be addressed by the Church, which, especially today, should be a reality pointing to the spiritual dimension of man. Therefore, this article will attempt to present the new challenges associated with the relationships that arise between humans and AI systems in the light of the Church's theological reflection. Initially, the article will illustrate certain natural predispositions of humans towards social life and relationality in the light of evolutionary and neurological sciences. Then, the development of artificial intelligence systems will be discussed, with special emphasis on the natural language processing by these systems, leading to increasingly intuitive and more “human” interaction with these technological creations. Finally, the theological perspective on assessing the possibilities and quality of relationships between humans and AI systems will be presented.

Keywords: technology, artificial intelligence, theology of communication

III. Sztuczna inteligencja w modelarstwie społecznym i teorii cybernetycznej

Copyright by Wydawnictwo Naukowe UKSW

Cybernetyczne modelarstwo społeczeństwa przyszłości w perspektywie narzędzia sztucznej inteligencji. ChatGPT a fantastyka i prognostyka socjologiczna

Wstęp

Socjologia – jak każda inna nauka – jest podatna na różne trendy i mody przejawiające się w akcentowaniu swoistych dla siebie obszarów poznawczych. Jednym z najbardziej popularnych jest obecnie zagadnienie sztucznej inteligencji. Coraz większa dostępność jej internetowych narzędzi (przede wszystkim programów opartych na formacie chatbotów) sprawia, że użytkownicy sieci używają ich niczym pewnych „podpowiadaczy”, a może nawet wyroczni w nurtujących ich kwestiach. Okazuje się, że podstawowa rola takich chatbotów, czyli konstruowanie odpowiedzi na pytania zawierające odwołania do zgromadzonych wielkich zasobów wiedzy (tekstów i obliczeń – głównie w formacie *big data*), może stać się przydatna również dla socjologów, którzy zastanawiają się nad kształtem i funkcjonowaniem ogólnie pojętego społeczeństwa przyszłości. Wprowadzanie chatbotów zmusza także do ponownego

¹ Dr Marcin Choczyński, Instytut Nauk Socjologicznych UKSW w Warszawie.

zastanowienia się nad kwestią „cybernetycznego modelarstwa społecznego” – pomysłu futurystycznego czy też koncepcji fabularnej – zaproponowanych przez Stanisława Lema, rozważanych do tej pory jedynie w pracach science fiction, a teraz skonkretyzowanych coraz bardziej jako realne narzędzia technologiczne, wyrażone też w formule *transformation design*.

Idea cybernetycznego modelarstwa społecznego i jej emanacja w fantastyce socjologicznej

Poszukiwanie charakterystyk społeczeństwa przyszłości należy niewątpliwie rozpocząć od przywołanego w tytule niniejszego opracowania cybernetycznego modelarstwa. Termin ten wprowadził znany futurolog i pisarz Stanisław Lem, który w jednej ze swoich prac teoretycznych pisał m.in. o konstruowaniu fabuł powieści:

Należy stać się empirykiem – budowniczym światów po prostu konstruowalnych, podług całości naszej dobrej wiedzy możliwych, jako obarczonych przynajmniej pewną, niechby znikomą nawet, szansą urzeczywistnienia. Nasuwa się wówczas koncepcja wprowadzenia w to pole tematyczne – metod cybernetycznego modelarstwa. Różne typy zależności powstających międzyludzko czy rozmaite typy struktur zbioru społecznego można będzie kiedyś badać [...] i obserwować tworzące się w maszynie procesy samoorganizacyjne, tendencje socjostatyczne lub socjolityczne, aby dojść jakości stanów równowagi, szukać krytycznych punktów socjodynamicznych przejść, co formację od formacji oddzielają, analizować trajektorie takich wymodelowanych społeczeństw wieloaspektowo, poddawać podobne organizmy pseudo-społeczne działaniu różnych czynników zakłócających ich równowagę, wymyślać więc wynalazki i odkrycia nie istniejące, dociekać skutków ich upowszechniania itd.²

² S. Lem, *Fantastyka i futurologia*, t. 2, Wydawnictwo Literackie, Kraków 1970, s. 332.

Twórca *Bajek robotów* czy też *Dzienników gwiazdowych* podpowiada zatem – szczególnie pisarzom – jak ciekawie i atrakcyjnie dla czytelników kreować świat fantastyczny i społeczeństwo w nim funkcjonujące, uwzględniając perspektywę strukturalno-konstruktywistyczną.

Lem podaje tutaj całkiem sporo cech *stricte* socjologicznych, m.in. zależności (relacje) czy też specyficzną budowę pewnych społecznych całości, w tym ogólne zarysy strukturalne. Następnie precyzuje, że mogą to być wprost odniesienia do reguł stratyfikacyjnych (zróznicowania społecznego – głównie w obszarze nierówności), zasad ruchliwości (awansów i degradacji wertykalnych czy też horyzontalnych). Zakłada ujęcie systemowe, zarówno w optyce statyki (schematów, modeli), jak i dynamiki społecznej (pragmatyki, działań). Kreśli momenty przejścia, czyli odwołuje się do zmian społecznych i ich punktów krytycznych (rewolucji), postuluje także zwrócenie uwagi na ewoluowanie społecznych struktur, ich progresywnizm, morfogenezę czy też załamania. Na końcu odnosi się także do zagadnienia różnych „atraktorów”, będących często forpocztami zmian, głównie w znaczeniu koncepcji determinizmu technologicznego.

Cybernetyczne modelarstwo społeczne najpełniej rozwinęło się w subgatunku fantastyki naukowej, określanym zbiorczo jako fantastyka socjologiczna (*social fiction*); pierwszoplanowym motywem stały się w niej „zagadnienia społeczne, głównie te, które mogą się pojawić w bliższej lub dalszej przyszłości na skutek wprowadzania bądź to daleko idących udogodnień życiowych, bądź też socjotechnicznych mechanizmów pozwalających na skuteczne sterowanie społeczeństwem”³. Antycypacja społeczna pojawiła się także we wcześniejszej literaturze utopistycznej – zarówno w wymiarze klasycznym (np. znane dzieło Tomasza Morusa, które dało nazwę temu obszarowi), jak i w słynnych XX-wiecznych antyutopiach, do których zalicza się powieści *My Jewgienija Zamiatina*, *Nowy wspaniały świat* Aldousa Huxleya czy też *Rok 1984* George’a Orwella. Są to tylko wybrane przykłady bardzo rozbudowanego katalogu dzieł. Polscy pisarze również bardzo aktywnie współtworzyli fantastykę socjologiczną – żeby przywołać tutaj wspomnianego już Stanisława Lema – ale należy także odnotować znane prace Janusza

³ A. Niewiadowski, A. Smuszkiewicz, *Leksykon polskiej literatury fantastycznonaukowej*, Wydawnictwo Poznańskie, Poznań 1990, s. 268.

A. Zajdla, Marka Oramusa czy też Edmunda Wnuka-Lipińskiego, *notabene* socjologa akademickiego⁴.

Modelarstwo przyszłości jest uwarunkowane pewną konwencją zastosowaną w literaturze *social fiction* – nowe społeczne światy są strukturalizowane przez pewien, najczęściej zewnętrzny, podmiot (może to być technologia wymykająca się spod kontroli, siła pozaziemska, uzurpatorska partia bądź ruch, ale też jakaś inna wyewoluowana rasa). Do tego zestawu obiektów kontrolujących reguły stratyfikacyjne dołączyła niedawno AI, dopełniając w zasadzie ową cybernetyczność – jako główną siłę transformacji. I to właśnie ona stała się odzwierciedleniem niepokojów związanych z brakiem wpływu człowieka na technologię, którą sam stworzył. Nie jest to jednak zupełnie nowy typ zagrożeń – wcześniej rolę takiego „straszaka” odgrywało chociażby np. klonowanie. Można zatem powiedzieć, że rzeczywistość „dogoniła” fabuły fantastyki socjologicznej (szczególnie tej sprzed kilkudziesięciu już lat), a AI może się okazać narzędziem, które to modelarstwo rzeczywiście wprowadzi, np. przez transhumanizm i przewyżczenie biologicznych oznak słabości ludzi⁵. Zagadnienie to domaga się jednak weryfikacji empirycznej.

Blaski i cienie naukowej prognostyki socjologicznej

Przyglądając się wszystkim wyróżnionym do tej pory w historii świata głównym modelom społeczeństwa, socjologowie nauczyli się wyodrębniać ich cechy i główne elementy strukturalne. Poszczególne typy społeczeństw: pierwotne (w tym zbieracko-łowieckie, kopieniackie, hortykulturowe i neolityczne), następnie wielkie cywilizacje przedindustrialne, społeczeństwa feudalne (tradycyjne), nowoczesne i ponowoczesne, zbudowane są na kanwie formacji społeczno-ekonomicznych, takich jak m.in. partykularyzm hord

⁴ M. Choczyński, *Jak zachęcić studentów socjologii do antycypacji społecznej i prognostyki societas futura? Udział fantastyki socjologicznej w dydaktyce akademickiej*, „Ruch Prawniczy, Ekonomiczny i Socjologiczny” 2023, t. 85, nr 2, s. 256–258.

⁵ S. Schneider, *Świadome maszyny. Sztuczna inteligencja i projektowanie umysłów*, tłum. J. Bednarek, PWN, Warszawa 2021, s. 25.

pierwotnych przeobrażający się w holizm i agraryzm u progu rewolucji neolitycznej, następnie imperializm doby wielkich cywilizacji przedindustrialnych, średniowieczny feudalizm, a także współczesne postaci kapitalizmu oraz socjalizmu i komunizmu. Do tego podstawowego rozróżnienia na pewno należy dodać systemy stratyfikacyjne, które określały antagonizmy i gradacje poszczególnych elementów struktur społecznych – od najwcześniejszych naturalnych (np. wyodrębnienie się zbiorowości pierwszych rolników z dotychczasowych hord zbieraczy-łowców i pasterzy), przez niewolnicze i kastowe (np. panowie i niewolnicy) w rozwijających się cywilizacjach, dalej tradycyjne stanowe (np. seniorzy – feudałowie i wasale – lennicy) i nowoczesne klasowo-warstwowe (np. burżuazja i proletariat). W systemach stratyfikacyjnych akcentowano czynniki (kapitały), wedle których ludzie się różnicowali, takie jak: środowisko, ziemia, pochodzenie czy finalnie zasoby materialne i intelekt. To z kolei wpływało na przypisany lub osiągniany charakter statusów społecznych⁶. Podane charakterystyki modernizacyjne powinny zatem być pewnymi wskazówkami dla tych badaczy, którzy próbują określić charakter zmian i przybliżyć założenia społeczeństwa tego typu, który pojawi się po epoce funkcjonującej obecnie, czyli ponowoczesności.

Kluczowym dla prognostyki socjologicznej pytaniem jest z pewnością to odnoszące się do nowego typu społeczeństwa, które nastąpi po ponowoczesności: w jaki sposób zostaną w nim ustanowione determinanty strukturalne czy też instytucjonalne? Jednak do tej pory – na polu naukowym – nie udało się odpowiedzieć na te pytania wprost. Istnieje oczywiście wiele dysput o końcu kapitalizmu jako systemu społeczno-ekonomicznego oraz wyczerpaniu dotychczasowego podziału klasowo-warstwowego, który w kapitalizmie jest naczelną formą stratyfikacji (zróżnicowania). Dużo miejsca zajmują analizy o kapitale (zasobie) jako takim i jego komasacji⁷. Oprócz tego wielu socjologów upatruje transformacji społecznych w przemianach idei, wynikających z nich ideologii oraz morfogenezie tradycyjnych instytucji

⁶ K. Krzysztofek, M. S. Szczepański, *Zrozumieć rozwój. Od społeczeństw tradycyjnych do informacyjnych. Podręcznik socjologii rozwoju społecznego dla studentów socjologii, nauk politycznych i ekonomii*, wyd. 2, Wydawnictwo UŚ, Katowice 2005, s. 35–37; 172–176.

⁷ T. Piketty, *Kapitał w XXI wieku*, tłum. A. Bilik, Wydawnictwo Krytyki Politycznej, Warszawa 2015, s. 141–144; G. Reisman, *Kapitał i kapitalizm XXI wieku czyli od błędnej teorii do destrukcyjnych reform Piketty'ego*, tłum. P. Nowakowski, Fijorr Publishing, Warszawa 2015, s. 33–37.

społecznych w kluczowych obszarach religii, rodziny, socjalizacji, ekonomii oraz polityki⁸, zapowiadających nadejście „nowego człowieka”.

Zasadniczą przeszkodą dla spójnego i zarazem trafnego przewidywania przyszłości społecznej w ramach socjologii jest problematyczna synteza determinizmu technologicznego ze społecznym konstruktywizmem⁹. Powoduje to, że antycypacja taka staje się równaniem z dwiema niewiadomymi – z jednej strony trudne jest dokładne oszacowanie rozwoju możliwości technologicznego rozwoju, a z drugiej – refleksji, co z taką futurystyczną technologią zrobią przyszłe generacje, jaki nadadzą jej sens czy też status. Wiele prognoz futurologicznych właśnie przez takie uwarunkowanie się nie sprawdza, co będzie widać na konkretnych przykładach.

Nie znaczy to jednak, że zagadnienia prognostyczne nie są formułowane w ogóle. Można je z grubsza podzielić na kilka obszarów interdyscyplinarnych. Najbardziej obszernym jest oczywiście katalog dzieł odwołujących się do postępującej dekadencji – czyli „końca” w wielu obszarach społecznych relacji. Dekadencja może się łączyć z ogólną niedoskonałością systemową, niewydolnością władz (ochłokracją) czy też kryzysami kulturowymi¹⁰. Oprócz tego „schyłkowość” jako główny temat odniesień widać w pracach socjoekonomicznych (również w marksizmie) oraz wywodzących się z nich koncepcjach końca ludzkiej pracy.

Warto na początku odnieść się do zmian w ekonomii, gdzie głównym motywem jest postępujący niedostatek zasobów, z którym będą musiały się mierzyć społeczeństwa przyszłości. Za przykład tego typu prognoz może służyć klasyczny maltuzjanizm¹¹ i jego kontynuacje (np. słynny raport Klubu Rzymskiego zatytułowany *Granice wzrostu*¹²). Innym przykładem zaadaptowania wyliczeń ekonometrycznych do socjologii jest chociażby koncepcja Simona Kuzneta, zakładająca spłaszczenie nierówności stratyfikacyjnych

⁸ S. Sowiński, *Wstęp*, w: *Co po postindustrializmie?*, red. K. Wielecki, S. Sowiński, Centrum Europejskie UW, Warszawa, s. 7–8.

⁹ K. Krzysztofek, *Prévoir – savoir – pouvoir czyli od przewidywania do wiedzy i władzy*, „Stan Rzeczy” 2018, nr 1(14), s. 29.

¹⁰ F. Znaniecki, *Upadek cywilizacji zachodniej. Szkic z pogranicza filozofii kultury i socjologii*, Gebethner i Wolff, Poznań 1921, s. 71.

¹¹ R. Bartkowiak, *Rozwój ludności świata i prawa nim rządzące*, SGH, Warszawa 2001, s. 28–31.

¹² Z. Sadowski, *Granice wzrostu – nowa ocena sytuacji*, „Przyszłość” 2006, nr 1, s. 11–14.

pod wpływem osiągnięcia wysokiego stopnia rozwoju społecznego¹³. Jak widać obecnie, żadna z tych prognoz się nie sprawdza – ludzkości potrzeba coraz więcej energii, nadal jest ona wytwarzana z paliw kopalnych, ale jest to proces daleko bardziej efektywny i coraz lepiej zdywersyfikowany, ponadto złoża surowców jak na razie się nie wyczerpują, za to odkrywane są wciąż nowe. Wzrasta świadomość ekologiczna i rozwijają się nowe technologie, które nie były do pomyślenia jeszcze 50 lat temu. Podobnie nie przewidział zintensyfikowania produkcji rolnej Thomas R. Malthus, wieszcząc na przełomie XVIII i XIX w. niemożność wyżywienia coraz większej liczby ludności. Ponadto współczesny kapitalizm raczej utrwała nierówności, niż je koryguje, co widać po rozkładzie światowych zasobów – są one rozproszone i stanowią własność nielicznych elit o różnym charakterze (często uzurpatorskim).

Bezpośrednio z teorii ekonomicznych swój początek biorą również te, które na swój sposób starają się oddać schyłkowość kapitalizmu, jako systemu słabnącego i zdegenerowanego – wśród nich prym wiedzie teoria światosystemów (systemów-światów) autorstwa Immanuela Wallersteina, opierająca się na uniwersalnym założeniu systemowym, z mechanizmami zarówno scalającymi, jak i rozprężającymi jego spoiwa, tworząc centra, peryferie i półperyferie¹⁴. Można też zaakcentować podobne do niej, jak te autorstwa członków tzw. szkoły dependystycznej, przeciwników kapitalizmu w jego obecnym kształcie¹⁵. Nie może tutaj też zabraknąć przewidywań samego Karola Marksa i jego kontynuatorów, wszak komunizm jako docelowy system stratyfikacyjny – oparty na bezklasowości – i jednocześnie powstałe na takim gruncie idealne społeczeństwo egalitarne miały stać się właśnie finalnym etapem rozwoju, który wprowadziłby tak pożądaną przez wielu filozofów i utopistów rzeczywistą równość społeczną, jednocześnie doszczętnie likwidując kapitalizm naznaczony permanentnymi kryzysami, wpisanymi immanentnie w jego istotę¹⁶.

¹³ P. Kumor, *Zależność nierówności płac od poziomu rozwoju gospodarczego*, „Gospodarka Narodowa” 2010, nr 7/8, s. 46–50.

¹⁴ E. Wallerstein, *Analiza systemów-światów. Wprowadzenie*, tłum. K. Gawlicz, M. Starnawski, Dialog, Warszawa 2007, s.10.

¹⁵ S. Amin, *Zmurszały kapitalizm*, tłum. R. Wojna, Z. M. Kowalewski, Dialog, Warszawa 2004, s. 118–121.

¹⁶ K. Świrek, *Karol Marks i problem przyszłości*, „Stan Rzeczy” 2018, nr 1(14), s. 209.

Pod koniec lat 80. XX w. istotnym echem odbiła się w socjologii koncepcja końca pracy autorstwa Jeremy'ego Rifkina¹⁷, który twierdził, że właśnie praca najemna zostanie w bliższej przyszłości zrobotyzowana tak, że człowiek nie będzie już w stanie jej wykonać, szczególnie w obszarze przemysłu¹⁸. Współczesne doświadczenia globalne nie potwierdzają jednak do końca tych założeń, gdyż informatyzacja i automatyzacja co prawda rzeczywiście likwidują pewne miejsca pracy, ale uruchamiają jednocześnie szereg innych, zatem nie jest to ruch skierowany – jak do tej pory – wyłącznie ku likwidacji pracy jako takiej. Chociaż z drugiej strony pojawiają się też głosy, że AI „będzie prowadzić do sytuacji, w której maszyny zaczną zastępować ludzi w myśleniu na wszystkich poziomach operacyjnych”¹⁹. Znacznie ciekawsze w polu widzenia przemian pracy jako środka transformacji społeczeństw wydają się koncepcje prekariatu, salariat²⁰ lub też innych pochodnych współczesnego proletariatu, np. cogitariat i digitariat, ukazujące strukturalne zmiany społeczne związane z istotą i jakością płac²¹.

Wyjściem poza ten ekonomiczny determinizm okazały się koncepcje prognozujące stały rozwój społeczeństwa sieci, związany z ekspansją internetu i niejako „zawładnięciem” społeczeństw przez wirtual i jego możliwości kreacyjne praktycznie w każdej dziedzinie. Globalny informacjonizm stał się zatem podstawą funkcjonowania człowieka²². Rozwinięciem idei sieciowości są z kolei koncepcje progresu społeczeństwa wiedzy powiązanego z tzw. dataizmem, czyli „władzą” danych na temat świata i jego istoty. Dataizm

¹⁷ J. Rifkin, *Koniec pracy. Schyłek siły roboczej na świecie i początek ery postrynkowej*, tłum. E. Kania, Wydawnictwo Dolnośląskie, Wrocław 2001, s. 17–21.

¹⁸ D. Walczak-Duraj, *Praca a nowe technologie. W kierunku transgresji etycznej i pracy nieodpłatnej w gospodarce cyfrowej*, „Przegląd Socjologiczny” 2019, t. 68, nr 3, s. 81–84; G. Szulczewski, *W poszukiwaniu sensu życia w świecie bez pracy*, w: *Świat (bez) pracy. Od fordyzmu do czwartej rewolucji przemysłowej. Monografia dedykowana profesorowi Juliuszowi Gardawskiemu*, red. J. Czarasty, Cz. Kliszko, SGH, Warszawa 2018, s. 32–39.

¹⁹ J. C. Lennox, *2084. Sztuczna inteligencja i przyszłość ludzkości*, tłum. Z. Kościuk, Prodo, Warszawa, 2023, s. 52.

²⁰ G. Standing, *Prekariat. Nowa niebezpieczna klasa*, tłum. K. Czarnecki, P. Kaczmarski, M. Karolak, PWN, Warszawa 2014, s. 43–53.

²¹ I.S. Fiut, M. Habryń, *Spółeczność akademicka wobec możliwości wykorzystania Internetu w procesie pracy*, w: *Mikrosocjeczność informacyjna na przykładzie miasteczka internetowego Akademii Górniczo-Hutniczej w Krakowie*, red. L. Haber, AGH, Kraków 2001, s. 162.

²² K. Korab, *Filozofia i socjologia wirtualnej rzeczywistości*, w: *Wirtual. Czy nowy wspaniały świat?*, red. K. Korab, Scholar, Warszawa 2010, s. 21.

traktowany jest jako synteza wszelkich osiągnięć cywilizacyjnych, które muszą być oparte na permanentnym gromadzeniu, kwantyfikacji i analizie danych (tzw. *big data*). W zamierzeniach zwolenników tej tezy takie działanie powinno przynieść nowy etap cywilizacyjny – oparty wyłącznie na wysoko rozwiniętej wiedzy²³. W najbardziej fantastycznym scenariuszu widać powstanie hybryd ludzi oraz maszyn i kreację właśnie takiego cybernetycznego społeczeństwa powiązanego wielostronnie ze sztuczną inteligencją²⁴. Byłby to zarazem rzeczywisty zwrot ku cybernetycznemu modelarstwu, gdyż pozwalałaby już na to technologia, dająca sobie radę z takim wyzwaniem.

Należy tutaj też wspomnieć o całej gamie prognoz futurologicznych (nazywanych łącznie *future studies*) – mniej lub bardziej ciekawych, łączących wiele stanowisk. Wśród nich trzeba na pewno wyróżnić znaną w socjologii koncepcję trzeciej fali²⁵ czy też polską szkołę futurologiczną²⁶. Co ciekawe, badacze wciąż poszukują nadchodzącej „czwartej fali” – być może rzeczywiście będzie to taka sztuczna inteligencja, która będzie potrafiła przenieść dotychczasowe osiągnięcia intelektualne człowieka w nowy wymiar i pokonać granice dotychczas nieprzekraczalne.

Jeśli uwzględnimy podane na wstępie założenia o technologicznym determinizmie i społecznym konstruktywizmie, wówczas przegląd tych koncepcji prowadzi do konstatacji o rzeczywistej nieprzewidywalności społecznej. Socjologia zatem w obecnym stanie wiedzy nie rysuje jednoznacznych obrazów przyszłości, nie została ponadto w żaden przekonujący sposób podana – choćby w zarysie – koncepcja „trzeciej drogi”, wychodząca poza uwarunkowania stratyfikacyjne kapitalizmu i komunizmu – chociaż z drugiej strony pojawiają się koncepcje kapitalizmu zmodyfikowanego, np. rentyzm²⁷. Z podanego wolumenu największym stopniem doprecyzowania, czyli

²³ L. W. Zacher, *Transformacje społeczeństw od informacji do wiedzy. Interdyscyplinarne wykłady, wpływ techniki i globalizacji*, C.H. Beck, Warszawa 2007, s. 207–247.

²⁴ Y. N. Harari, *Sapiens. Od zwierząt do bogów*, tłum. J. Hunia, Wydawnictwo Literackie, Warszawa 2018, s. 66–67.

²⁵ A. Toffler, *Trzecia fala*, tłum. E. Woydyłło, M. Kłobukowski, Wydawnictwo Kurpisz, Poznań 2006, s. 167.

²⁶ L. Becht, P. Filipkowski, *Ćwiczenie wyobraźni socjologicznej: polska futurologia naukowa lat 70. XX wieku*, „Stan Rzeczy” 2018, nr 1(14), s. 43–59.

²⁷ P. Frase, *Cztery przyszłości. Wizje świata po kapitalizmie*, tłum. M. Szlinder, PWN, Warszawa 2018, s. 30; 62–78.

zakreślenia następnego etapu periodyzacji społeczeństw, wychodzącego poza współczesny kapitalizm, charakteryzuje się grupa teorii marksistowskich. Inne opierają się jedynie na przypuszczeniach i „gdybaniach”, nie dając jednak jednoznacznych odpowiedzi na pytanie o faktyczny strukturalizm *societas futura*. Brak konkretyzacji i rozmycie prognoz spowodowały, że faktyczne budowanie alternatywnych światów społecznych na długie lata pozostało domeną literatury *social fiction*, która jako jedyna mogła wykonać ten „krok naprzód” i wymodelować tak fantastyczny kształt społeczeństwa, na jaki tylko pozwalał talent i zamysł danego twórcy.

Zagadnienia przyszłości społecznej w kontekście AI

Sztuczna inteligencja, oznaczana różnymi skrótowcami (najczęściej AI – od angielskiego określenia *artificial intelligence*, a na gruncie języka polskiego jako SI bądź SzI), stała się w ostatnim czasie jednym z głównych tematów odniesień naukowych – przede wszystkim z uwagi na swoją nowatorskość i zaciekawienie jej rezultatami. Oprócz oczywistych odniesień do nauk informatycznych i ogólnie przyrodniczych AI została implementowana także w obszar humanistyki, czy też językoznawstwa, jako narzędzie przydatne w synteźowaniu tekstów kultury i opracowań *stricte* naukowych. O popularności tematyki AI wśród obecnych badaczy i autorów świadczy „wysyp” wielu pozycji wydawniczych jej poświęconych – przekrój formalny tych dzieł jest bardzo szeroki, od syntez naukowych aż po beletrystykę, którą można nazwać nowym rozdaniem lub też nową, interesującą konwencją, zaistniałą w science fiction, a ściślej rzecz biorąc – w fantastyce socjologicznej.

Pierwszą definicję sztucznej inteligencji zaproponował John McCarthy już w 1956 r. Rozumiano ją wtedy jako „zdolność systemu do prawidłowego interpretowania danych pochodzących z zewnętrznych źródeł, nauki na ich podstawie oraz wykorzystywania tej wiedzy, aby wykonać określone zadania i osiągnąć cele poprzez elastyczne dostosowanie”²⁸. Kluczowa jest tutaj zdolność adaptacji do zastanych warunków, ale też czerpanie maksimum

²⁸ F. Kurp, *Sztuczna inteligencja od podstaw*, Helion, Gliwice 2023, s. 11.

użytecznej dla systemu wiedzy ze środowiska – w tym przypadku zbiorczego środowiska życia człowieka, czyli ze społeczeństwa.

Za kreację i udoskonalanie AI odpowiada przede wszystkim człowiek, który stara się w jakimś stopniu zapanować nad kulturą, ujmowaną jako wielki zbiór danych wytworzonych przez samych ludzi. Makroskalowy charakter tych danych domagał się wprowadzenia takiego rozwiązania, które szybko i skutecznie rozwiązywałoby problemy interpretacji czy też syntezy. Dlatego też:

[...] sztuczna inteligencja ma dwa główne cele. Pierwszy ma charakter **technologiczny** – sprawienie, by komputery robiły pożyteczne rzeczy (często przy wykorzystaniu metod mających **niewiele** wspólnego z działaniem umysłów). Drugi ma charakter **naukowy** – polega na wykorzystaniu pojęć i modeli sztucznej inteligencji w taki sposób, by mogły one odpowiedzieć na pytania dotyczące ludzi i innych istot żywych²⁹.

Tendencje do poznawania mozaiki ludzkiej wiedzy stały się tak naglące z uwagi na ogromny przyrost idei, rozwiązań i samych wynalazków, które wpływając wzajemnie na siebie, określają globalne ścieżki rozwoju.

Warto w tym miejscu się zastanowić, jak konkretnie rozwój AI wpływa na społeczne (zbiorowe) determinanty ludzkiego funkcjonowania. Przede wszystkim akcent położony jest tutaj na fakt implementacji rozwiązań sztucznej inteligencji w społeczeństwie, dopuszczenie do współistnienia tej technologii wraz z ludźmi.

Akceptacja regulacji normatywnej społeczeństwa z udziałem SI jako członka tego społeczeństwa zakładałaby bowiem podstawową tezę, że społeczeństwo to nie jest już w sensie czystym „społeczeństwem ludzkim”. Staje się ono zupełnie nowym tworem socjologicznym, przy czym socjologia, psychologia czy inne nauki społeczne mają do wyboru albo się rozwarstwić na nauki ukierunkowane na SI (ale czy tylko przez SI prowadzone) i na nauki ukierunkowane na człowieka, lub też – na konglomerat społeczeństwa, którego członkami są zarówno ludzie, jak i SI. To, czy ludzie zaakceptują

²⁹ M. A. Boden, *Sztuczna inteligencja. Jej natura i przyszłość*, tłum. T. Sieczkowski, seria „Krótkie Wprowadzenie” UŁ, Łódź 2020, s. 14.

społeczeństwo będące konglomeratem ludzi i SI, jest pytaniem równie zasadnym, jak to, czy będą mieli inne wyjście. I to, czy faktycznie pozostaną w takim społeczeństwie równouprawnionymi członkami³⁰.

W jednej z prób zdefiniowania roli sztucznej inteligencji w odniesieniu do kapitału Maria Chilicka stwierdziła, że „SI to forma automatyzacji, a automatyzacja to równoważnik kapitału. Korzystają wyłącznie jego właściciele. SI przechyla szalę na korzyść kapitału, a nie klasy robotniczej”³¹. Stosowanie sztucznej inteligencji nie pozostaje zatem bez wpływu na globalne uwarunkowania napięć między klasami zarządzającymi a wykonawczymi.

Przyglądając się rozwojowi konkretnych typów społeczeństw w perspektywie technologicznej, można wyróżnić szereg tych rzeczy (ściślej rzecz biorąc – wynalazków), które okazywały się prawdziwymi determinantami rozwoju, ukierunkowując wręcz ludzi na nowe formy pracy, gospodarowania, planowania czy nawet stylów życia. W literaturze przedmiotu można zaobserwować stanowisko, że tymi rzeczywiście istotnymi dla zmiany społecznej wynalazkami były maszyna parowa, elektryczność i komputery. Te mają już znaczenie historyczne, natomiast współcześnie za taki wynalazek uchodzi właśnie sztuczna inteligencja³².

Kontynuując ten wątek, można powiązać wskazane powyżej kluczowe wynalazki z podziałem na charakterystyczne dla danego czasu formy przemysłu. I tak w XVIII w. wyewoluował tzw. przemysł 1.0 – oparty na mechanizacji (a głównie maszynie parowej), co zresztą zapoczątkowało samą industrializację *sensu stricto*. Drugim etapem – czyli przemysłem 2.0 – była ta forma produkcji, w której maszyny parowe zastępowała w XIX w. elektryfikacja, co pozwoliło z kolei na wielkoseryjność produkcji (uzewnętrzną chociażby w idei fordyzmu). Z kolei w latach 70. XX w. zaczęto wprowadzać cyfryzację i automatyzację do sterowania maszynami, określając to jako przemysł 3.0, w ramach którego oprogramowanie zaczęło zastępować całe

³⁰ A. Chłopecki, *Sztuczna inteligencja – szkice prawne i futurologiczne*, wyd. 2, C.H. Beck, Warszawa 2021, s. 130.

³¹ M. Chilicka, *Sztuczna inteligencja. SI – to już nie science fiction*, Temat, Bydgoszcz 2020, s. 56.

³² Tamże, s. 47.

rzezse miejskiego proletariatu. Ostatni jak do tej pory wyróżniony etap, czyli przemysł 4.0, zakłada połączenie produkcji, automatyzacji i technik wytwórczych pod egidą wieloskładnikowej analizy danych. Można tu wymienić m.in. systemy cyberfizyczne, internet rzeczy oraz przetwarzanie chmurowe³³. Jak widać, tak wyrażony determinizm technologiczny jest odpowiedzialny zarówno za kreację wielkich zbiorów ludzkich (klas społecznych), jak też ich postępującą deprecjację.

W ślad za tym zaczęto formułować paradygmaty sztucznej inteligencji związane z możliwościami systemów, na których ona się opiera. Do pierwszego paradygmatu zaliczyć można „symboliczną SI – składającą się między innymi z indukcyjnego programowania logicznego, automatyzacji procesów robotycznych i systemów eksperckich (w tym systemów rozmytych)”³⁴. Bardziej złożona technologia przetwarzania danych wymusiła powstanie sztucznej inteligencji, którą zaczęto nazywać „statystyczną SI – składającą się z sieci decyzyjnych, programowania probabilistycznego (w tym syntezy programów bayesowskich), widzenia komputerowego (działania i rozpoznawania obrazu, a także wizji maszynowej), przetwarzania języka naturalnego i uczenia maszynowego”³⁵. Ostatnim paradygmatem sztucznej inteligencji, akcentującym jej najbardziej złożony wymiar, jest stan rzeczy określony „symboliczną SI – polegającą na rozproszonej sztucznej inteligencji (w tym modelowaniu opartym na agencie, inteligencji roju i systemach wieloagentowych), obliczeniach afektywnych, inteligencji ucieleśnionej i różnych systemach autonomicznych”³⁶. Widać zatem postępujące zbliżanie się sztucznej inteligencji do myślenia ludzkiego, czyli wiązania ze sobą różnicujących składowych, których efektem jest przede wszystkim myślenie nieszablonowe oraz kreatywne.

Kolejnym definicyjnym rozwiązaniem w tym zakresie było przyjęcie podziału sztucznej inteligencji na słabą (*weak AI*) i silną (*strong AI*). Pojęcia te zostały wprowadzone przez Johna Searle’a w pracy *Minds, brains and*

³³ K. Grochowina, *Sztuczna inteligencja fundamentem innowacji w życiu i w biznesie*, w: *Świat nowych technologii. Czy sztuczna inteligencja zdominuje życie człowieka?*, red. J. Stawnicka, D. Morańska, W. Kubies, Humanitas, Sosnowiec 2019, s. 110.

³⁴ A. Przegalińska, P. Oksanowicz, *Sztuczna inteligencja nieludzka, arcyłudzka. Fenomen świata nowych technologii*, Znak, Kraków 2023, s. 47.

³⁵ Tamże.

³⁶ Tamże.

programs (1980). Słaba sztuczna inteligencja jest wykorzystywana głównie do rejestrowania i rozpoznawania danych pochodzących z otoczenia (np. rozpoznawanie mowy, reakcja na algorytmy). Stosuje się ją w sposób problemowy – do rozwiązywania zagadnień i osiągania celów. Ważną cechą takiej AI jest zasadniczy brak świadomości, ale także zdolności do samoświadomego rozumowania. Jest ona zatem ograniczona jedynie do zakresu przypadków, do jakich została zaprojektowana³⁷.

Rozszerzeniem z kolei tych rozwiązań elementarnych jest tzw. silna sztuczna inteligencja, która w przeciwieństwie do swojej starszej i jednocześnie mniej wydajnej wersji ma już zdolność samodzielnej kreacji rozwiązań, ale też współuczestnictwa z człowiekiem w praktycznie wszystkich zadaniach intelektualnych. Dodatkowo silna sztuczna inteligencja „myśli” dokładnie tak samo, jak człowiek, dostosowując się twórczo do zastanych sytuacji i wykorzystując adaptacyjnie dane warunki środowiskowe³⁸.

Współcześnie wiele dyskusji wywołują społeczne eksperymenty związane z coraz częstszym rzeczywistym stosowaniem idei AI do kontroli obywateli. Można nawet powiedzieć, że zagadnienia inwigilacji i dostępu do informacji definiujących człowieka (w tym przeważnie tych wrażliwych) stają się również formą kapitału, który ma przełożenie na zysk, stąd też jest wdrażany. Może to się odbywać zarówno przez przedsiębiorstwa komercyjne, jak i władze państwowe – system taki określa się w literaturze jako kapitalizm nadzoru, gdzie obywatele (a może już zbywatele?) poddani są ciągłemu sprawdzaniu i inwigilacji³⁹, głównie przez monetyzację wyborów, doznań i doświadczeń.

Przykładem tego drugiego obszaru jest na pewno nowy ustrój społeczno-polityczny, znany jako algokracja i wprowadzany współcześnie w Chińskiej Republice Ludowej. Jest to system oparty na powszechnym scoringu społecznym – tzw. obliczaniu kredytu obywateli (czyli *social score*). Przy czym pojęcie kredytu ma tutaj nieco inne znaczenie niż w przypadku rozumienia europejskiego czy też amerykańskiego. Kredyt bowiem w tradycji konfucjańskiej

³⁷ T. Muras, G. Margas, *ChatGPT-4. Sztuczna inteligencja. Czym jest i jak może zmienić nasz świat*, Fundacja „Zmieniam”, Szczecin 2023, s. 39.

³⁸ Tamże, s. 39.

³⁹ S. Zuboff, *Wiek kapitalizmu inwigilacji. Walka o przyszłość ludzkości na nowej granicy władzy*, tłum. A. Unterschuetz, Zysk i S-ka, Poznań 2020, s. 9; M. Fleischer, *Kapitalizm i jego sztuczna inteligencja*, LIBRON – Filip Lohner, Kraków 2022, s. 38.

jest zbliżony do idei kapitału bądź zasobu, gdyż oznacza „aktywa materialne i niematerialne, które posiada dana jednostka lub rządzący”⁴⁰. Na plan pierwszy wysuwa się tutaj zatem poznanie struktury informacyjnej, czyli pewnego zbioru danych o obywatelu, i określenie jego charakterystyki ekonomicznej, światopoglądowej i zachowaniowej. „Należy zbierać dane o jednostkach, firmach i organizacjach, zaś zakres zbieranych danych ma być bardzo szeroki, obejmując dane finansowe, zakupy w e-commerce, korzystanie ze środków transportu czy aktywność sportową”⁴¹. Można powiedzieć, że podobnie jak oblicza się zdolność kredytową, tak tutaj oblicza się zdolność społeczną, czyli po prostu użyteczność jednostki dla państwa czy też jego konkretnych instytucji. Algokracja jest więc pewnym pomysłem na uporządkowanie sfery funkcjonowania człowieka, a współczesne możliwości w analizie *big data* umożliwiają technologicznie śledzenie wręcz każdego kroku człowieka i całości jego egzystencji. Stanowi ona też pewną kontynuację konceptu netokracji, zaakcentowanego kilkanaście lat wcześniej⁴².

Rozmowy z ChatGPT o antycypacji społecznej

W dniu 20 października 2023 r. autor niniejszego tekstu zadał programowi ChatGPT w wersji 3.5 cztery pytania odnośnie do antycypacji społecznej, zawierające różne poziomy konkretyzacji. Autor wiązał z tą operacją duże nadzieje – sądził, że AI pozwoli mu skierować ciekawość poznawczą na nowe tropy oraz ukazać nieoczywiste do tej pory nawiązania. Wynikało to po części z cechy AI, jaką jest podanie opcji z dowolnej liczby teoretycznych i praktycznych dowodów⁴³. Zarówno pytania, jak i odpowiedzi można prześledzić na poniższych czterech zrzutach ekranu, ukazujących efekt pracy tego konkretnego narzędzia AI.

⁴⁰ K. Rybiński, J. Królewski, *Algokracja. Jak i dlaczego sztuczna inteligencja zmienia wszystko?*, PWN, Warszawa 2023, s. 143.

⁴¹ Tamże, s. 144.

⁴² A. Bard, J. Söderqvist, *Netokracja. Nowa elita władzy i życie po kapitalizmie*, tłum. P. Cypriański, Wydawnictwa Akademickie i Profesjonalne, Warszawa 2006, s. 41.

⁴³ J. Kaplan, *Sztuczna inteligencja. Co każdy powinien wiedzieć*, tłum. S. Szymański, PWN, Warszawa 2019, s. 18.

Obraz 1.

Default (GPT-3.5)

MA

Sztuczna inteligencjo, napisz mi proszę, jak będzie wyglądało społeczeństwo przyszłości?



Przewidywanie, jak dokładnie będzie wyglądać społeczeństwo przyszłości, jest trudne i pełne niepewności. Jednak można wskazać kilka ogólnych trendów i obszarów, które mogą mieć wpływ na kształtowanie społeczeństwa w przyszłości. Poniżej przedstawiam kilka kluczowych kwestii, które mogą wpłynąć na to, jak będzie wyglądało społeczeństwo przyszłości:



1. **Technologia i sztuczna inteligencja:** Rozwój technologii, w tym sztucznej inteligencji, będzie miał ogromny wpływ na społeczeństwo przyszłości. Automatyzacja pracy, rozwijające się systemy opieki zdrowotnej oparte na AI, autonomiczne pojazdy i wiele innych innowacji mogą znacząco zmienić nasze życie.
2. **Przemysł 4.0:** Rozwinięcie przemysłu 4.0, który opiera się na cyfryzacji i automatyzacji produkcji, może zmienić model ekonomiczny i strukturę zatrudnienia. To może oznaczać, że niektóre tradycyjne zawody stracą na znaczeniu, a nowe pojawią się w odpowiedzi na rosnący popyt na umiejętności związane z technologią.
3. **Demografia:** Starzejące się społeczeństwo i zmiany w strukturze wiekowej mogą wpłynąć na systemy emerytalne, opieki zdrowotnej i innych sektorów. Ponadto migracje i zmiany w składzie etnicznym mogą kształtować kulturę i społeczeństwa.
4. **Zrównoważony rozwój:** W miarę narastającego problemu zmian klimatycznych i zanieczyszczenia środowiska, społeczeństwo przyszłości będzie musiało bardziej niż kiedykolwiek zwracać uwagę na kwestie zrównoważonego rozwoju i ochrony środowiska.
5. **Edukacja i umiejętności:** Zmiany w wymaganiach na rynku pracy będą wpływać na potrzeby edukacyjne. Rozwinięcie zdolności adaptacji, kreatywności i umiejętności cyfrowych stanie się kluczowe.
6. **Medycyna i długowieczność:** Postęp w medycynie i biotechnologii może przynieść zwiększoną długowieczność, co wpłynie na pytania związane z opieką zdrowotną, emeryturami i strukturą rodzin.
7. **Komunikacja i kultura:** Internet i media społecznościowe nadal będą odgrywać istotną rolę w komunikacji i kształtowaniu kultury. Mogą również wpływać na to, jak społeczeństwo przetwarza informacje i podejmuje decyzje.
8. **Etyka i prywatność:** W miarę jak technologia będzie się rozwijać, pojawiać się będą pytania dotyczące etyki i prywatności, a społeczeństwo będzie musiało rozważać, jak równoważyć korzyści technologiczne z ochroną prywatności i bezpieczeństwa.

Warto podkreślić, że przyszłość jest niepewna, i wiele czynników może wpłynąć na to, jak społeczeństwo będzie się rozwijać. Jednak wspomniane wyżej obszary stanowią istotne aspekty, które mogą mieć znaczący wpływ na kształtowanie społeczeństwa w przyszłości.

Obraz 2.

Default (GPT-3.5)

MA

Sztuczna Inteligencjo, napisz mi proszę jak zmieni się struktura społeczna w przyszłości? Czy nadal będziemy funkcjonować w systemie kapitalistycznym (klasowo-warstwowym), czy może jakimś innym? Uwzględnij wiedzę z socjologii i dotychczasowych zmian społecznych.



Przewidywanie dokładnych zmian w strukturze społecznej w przyszłości jest trudne, ale można rozważyć kilka kierunków ewolucji, uwzględniając istniejącą wiedzę z socjologii i dotychczasowe zmiany społeczne. Pamiętaj jednak, że to jedynie spekulacje, a rzeczywistość może być znacznie bardziej złożona i zmienna.

1. Ewolucja systemu ekonomicznego:
 - System kapitalistyczny jest obecnie dominującym modelem ekonomicznym na świecie. Jednak jego kształt i funkcjonowanie mogą ulec zmianie w przyszłości. W miarę jak technologia rozwija się, pojawić się może większa automatyzacja produkcji, co może wpłynąć na strukturę zatrudnienia i relacje pracodawca-pracownik. To może prowadzić do konieczności dostosowania systemu ekonomicznego, aby zapewnić sprawiedliwość społeczną i równość szans.
2. Wzrost znaczenia edukacji:
 - Wiedza i wykształcenie są coraz ważniejsze w gospodarce opartej na wiedzy. W przyszłości możemy oczekiwać, że inwestycje w edukację będą kluczowe. To może prowadzić do bardziej otwartego i elastycznego rynku pracy oraz bardziej zrównoważonych społeczeństw.
3. Wzrost znaczenia nowych technologii:
 - Rozwój technologii informatycznych, sztucznej inteligencji i automatyzacji będzie miał wpływ na strukturę społeczną. Będzie to miało zarówno pozytywne, jak i negatywne skutki. Może to zwiększyć nierówności w dochodach, ale również stworzyć nowe możliwości i zawody.

4. Wzrost wielokulturowości:
 - Globalizacja i migracje spowodują wzrost różnorodności kulturowej w społeczeństwach. To może prowadzić do bardziej otwartych i różnorodnych społeczeństw, ale może także rodzić wyzwania związane z integracją i tolerancją.
5. Zmiany demograficzne:
 - Starzenie się populacji w wielu krajach będzie miało wpływ na strukturę społeczną. To może prowadzić do konieczności dostosowania systemów opieki społecznej i emerytalnych.
6. Zrównoważony rozwój:
 - Rosnące zanieczyszczenie środowiska i zmiany klimatyczne będą miały wpływ na strukturę społeczną w przyszłości. Może to prowadzić do większego nacisku na zrównoważony rozwój i bardziej ekologiczne gospodarki.
7. Ewolucja struktury społecznej:
 - Struktura społeczna może ulec ewolucji w kierunku większej równości i elastyczności, z mniejszą liczbą tradycyjnych klas społecznych. Jednak zmiany te mogą być stopniowe i zależeć od wielu czynników.

Podsumowując, przyszła struktura społeczna będzie prawdopodobnie wynikiem oddziaływania wielu czynników, takich jak rozwijająca się technologia, zmiany demograficzne, ekonomiczne i kulturowe. System kapitalistyczny może ulec ewolucji, ale czy zostanie zastąpiony innym modelem, zależy od wielu czynników i trudno jest to jednoznacznie przewidzieć. Ważne jest, aby podejść do tych zmian z elastycznością i dążyć do tworzenia bardziej sprawiedliwego i zrównoważonego społeczeństwa.

Obraz 3.

Default (GPT-3.5)

MA

Sztuczna Inteligencjo, napisz mi proszę, co będzie w przyszłości decydowało o społecznym zróżnicowaniu (stratyfikacji)? Czy będzie to jakaś nowa forma kapitału (zasobu) czy może jakiś inny czynnik?



Przyszłe czynniki decydujące o społecznym zróżnicowaniu (stratyfikacji) są złożone i zależą od wielu czynników. Sztuczna inteligencja (SI) i technologie z nią związane na pewno będą miały wpływ na kształtowanie się tej stratyfikacji społecznej. Oto kilka możliwych czynników, które mogą odegrać rolę w przyszłym społecznym zróżnicowaniu:

1. **Kapitał cyfrowy:** W przyszłości, dostęp i umiejętność wykorzystywania technologii, w tym SI, mogą stać się nową formą kapitału. Osoby i firmy, które dysponują zaawansowanymi umiejętnościami w obszarze SI, mogą mieć przewagę konkurencyjną na rynku pracy i w biznesie. Jednocześnie, ci, którzy nie są w stanie nadążyć za postępem technologicznym, mogą być wykluczeni.
2. **Edukacja:** Wiedza i umiejętności związane z SI będą miały wpływ na społeczną stratyfikację. Osoby z dostępem do zaawansowanej edukacji z obszaru SI będą miały większe szanse na sukces zawodowy, podczas gdy osoby bez takiego dostępu mogą pozostać w tyle.
3. **Dostęp do technologii:** Dostęp do zaawansowanych technologii, w tym narzędzi opartych na SI, będzie istotnym czynnikiem wpływającym na jakość życia i możliwości zawodowe. Osoby, które nie mają dostępu do komputera, Internetu czy inteligentnych urządzeń, mogą być wykluczone z wielu aspektów życia społecznego.
4. **Praca i automatyzacja:** Automatyzacja pracy, częściowo dzięki SI, może prowadzić do zwiększenia nierówności dochodowych. Osoby o umiejętnościach, które nie są łatwo zautomatyzowane, mogą być mniej narażone na utratę pracy i związaną z tym stratę dochodów.
5. **Etyka i regulacje:** Jak społeczeństwo podejdzie do etycznych kwestii związanych z SI, takich jak prywatność, dyskryminacja czy bezpieczeństwo, będzie miało wpływ na to, jak technologia ta wpłynie na zróżnicowanie społeczne.
6. **Kultura i społeczeństwo:** Kwestie kulturowe, takie jak różnice w wartościach, przekonaniach i normach społecznych, nadal będą odgrywać kluczową rolę w społecznym zróżnicowaniu.

Podsumowując, przyszła stratyfikacja społeczna będzie zależała od wielu czynników, w tym rozwoju technologii, edukacji, dostępu do zasobów, pracy, etyki i kultury. Sztuczna inteligencja i technologie z nią związane będą jednym z kluczowych elementów wpływających na te kwestie, ale nie będą jedynym determinantem społecznego zróżnicowania.

Obraz 4.

Default (GPT-3.5)



Sztuczna Inteligencjo, napisz mi proszę, czy ty konkretnie będziesz modelować społeczeństwa przyszłości? Czy będziesz korzystać z jakiś znanych już systemów (np. feudalizm, kapitalizm, komunizm)? Czy będziesz modelować według konkretnych cech (np. uroda, robotyzacji)?



Jestem tylko programem komputerowym stworzonym do przetwarzania tekstu i udzielania odpowiedzi na pytania. Nie jestem w stanie modelować społeczeństw przyszłości ani korzystać z jakichkolwiek znanych systemów społeczno-ekonomicznych, takich jak feudalizm, kapitalizm czy komunizm. Moje możliwości ograniczają się do dostarczania informacji i sugestii na podstawie tekstu, który jest mi dostarczany.

Modelowanie społeczeństw przyszłości jest zadaniem skomplikowanym i wymaga interdyscyplinarnej wiedzy oraz analizy wielu czynników społecznych, ekonomicznych i technologicznych. To jest dziedzina badań, którą podejmują ludzie, tacy jak socjologowie, ekonomiści, politolodzy i inni eksperci.

Jeśli masz pytania dotyczące konkretnych cech społeczeństwa, takich jak uroda czy robotyzacja, mogę dostarczyć informacje na ten temat w oparciu o moją wiedzę do września 2021 roku. Jednak to ludzie, a nie programy komputerowe, podejmują decyzje dotyczące kształtu społeczeństwa i jego przyszłego rozwoju.

Konkluzja

Po lekturze odpowiedzi udzielonych przez sztuczną inteligencję widać wyraźnie, że ogólnodostępne darmowe i najbardziej rozpowszechnione narzędzie AI (ChatGPT ver. 3.5, dystrybutor: OpenAI) w dniu dzisiejszym nie potrafi (jeszcze!) skonkretyzować społeczeństwa przyszłości, ale podaje jego zasadnicze cechy, zgodne z wolumenem dostępnej literatury przedmiotu na temat prognostyki socjologicznej. Cechy te związane są przede wszystkim z szeroko rozumianym progresywizmem w wielu dziedzinach: medycyny, edukacji i automatyzacji pracowniczej. Można wszelako domniemywać, że zapewne tzw. wersje premium ChatGPT (np. kolejny, udoskonalony typ 4.0), jak też podobne programy od innych producentów (m.in. Bard – Google czy też HuggingChat – LAION) udzieliłyby odpowiedzi o nieco innym sprofilowaniu – jest to jednocześnie zachęta do kontynuacji badań w tym obszarze⁴⁴. Wydaje się więc, że ChatGPT 3.5 jest w obecnej sytuacji nie w pełni doskonałym narzędziem (aczkolwiek powszechnym i dostępnym bez płatnych licencji dla każdego użytkownika sieci), warto jednak dodać, że istnieją już nowsze wersje, które mogą zmienić wysunięte w niniejszym opracowaniu wnioski.

Charakter przytoczonych odpowiedzi na pytania i społecznej antycypacji ukazuje, że współczesna AI w postaci programu ChatGPT 3.5 znajduje się w tzw. fali percepcyjnej, czyli ma możliwości rozpoznawania i grupowania pewnych faktów⁴⁵, ale jednocześnie nie potrafi jeszcze tego twórczo zmodyfikować, co zakładała koncepcja fali autonomicznej (będącej połączeniem fali internetowej, biznesowej i percepcyjnej), czyli nie osiągnęła stanu rzeczy, jakim jest „zdolność maszyn do optymalizacji na podstawie złożonych zbiorów danych z nowo nabytym potencjałem sensorycznym”⁴⁶. Dopiero w czasie urzeczywistnienia się tej fali zakłada się rzeczywiste dojście do cybernetycznego modelarstwa społecznego, gdyż „maszyny łączące

⁴⁴ R. Hoffman, *GPT4, Rozmowa z Chatem GPT o przyszłości ludzi i świata*, tłum. K. Mirowski, Prześwity, Warszawa 2023, s. 123–127.

⁴⁵ K.-F. Lee, *Inteligencja sztuczna, rewolucja prawdziwa. USA, Chiny i przyszłość świata*, tłum. K. Hejnowski, Media Rodzina, Poznań 2019, s. 133–145.

⁴⁶ Tamże, s. 158.

w sobie te nadludzkie możliwości będą nie tylko rozumieć świat wokół nich – będą go także kształtować⁴⁷.

Przeprowadzona krótka analiza uświadomiła konieczność poszukiwania innych od ChatGPT narzędzi AI do prognostyki socjologicznej. Co istotne, ChatGPT w tym zestawieniu „odsuwa” od siebie możliwości kreatywne (czyli właśnie cybernetyczne modelarstwo) na rzecz ludzi – wiadać to w czwartej, ostatniej odpowiedzi. Ponadto należy zwrócić też uwagę na to, że ChatGPT 3.5 – zaprogramowany w formule dużego modelu językowego – funkcjonuje, opierając się na zasobach internetowych (zakres informacji dostępnych do analiz – wyłącznie do czerwca 2021 r.). Jego dane stanowią wynik tego, co ludzkość do tej pory stworzyła czy zaproponowała w sferze koncepcyjnej oraz praktycznej. ChatGPT 3.5 nie ma zatem pełnej zdolności kreowania nowych informacji, a jego prognostyka opiera się na ludzkich przewidywaniach jedynie zsumowanych przez program. Jest to więc w zasadzie symulacja składająca się z wypadkowej dostępnych „na wejściu” informacji. Kwestią otwartą jest natomiast pytanie, do czego będą zdolne nowe boty językowe? Jednak w tej dziedzinie postęp jest wykładniczy, zatem nie sposób przewidzieć, jak będzie się rozwijała AI właśnie w kontekście antycypacji społecznej. Najbliższa przyszłość pokaże więc, jakie będą dalsze udoskonalenia w tym obszarze. Można jednak stwierdzić, że najbardziej wysublimowane cybernetyczne modelarstwo społeczne zachodzi zdecydowanie częściej nadal w obszarze literackiej fantastyki socjologicznej i jej audiowizualnej adaptacji.

Bibliografia

- Amin S., *Zmurszały kapitalizm*, tłum. R. Wojna, Z.M. Kowalewski, Dialog, Warszawa 2004.
- Bard A., Söderqvist J., *Netokracja. Nowa elita władzy i życie po kapitalizmie*, tłum. P. Cypryański, Wydawnictwa Akademickie i Profesjonalne, Warszawa 2006.
- Bartkowiak R., *Rozwój ludności świata i prawa nim rządzące*, SGH, Warszawa 2001.
- Becht L., Filipkowski P., *Ćwiczenie wyobraźni socjologicznej: polska futurologia naukowa lat 70. XX wieku*, „Stan Rzeczy” 2018, nr 1(14), s. 41–69.

⁴⁷ Tamże, s. 158.

- Boden M. A., *Sztuczna inteligencja. Jej natura i przyszłość*, tłum. T. Sieczkowski, „Krótkie Wprowadzenie” 21, UŁ, Łódź 2020.
- Chilicka M., *Sztuczna inteligencja. SI – to już nie science fiction*, Temat, Bydgoszcz 2020.
- Chłopecki A., *Sztuczna inteligencja – szkice prawne i futurologiczne*, wyd. 2, C.H. Beck, Warszawa 2021.
- Choczyński M., *Jak zachęcić studentów socjologii do antycypacji społecznej i prognozyki societies futura? Udział fantastyki socjologicznej w dydaktyce akademickiej*, „Ruch Prawniczy, Ekonomiczny i Socjologiczny” 2023, t. 85, nr 2, s. 253–267.
- Fiut I. S., Habryń M., *Społeczność akademicka wobec możliwości wykorzystania Internetu w procesie pracy*, w: *Mikrospołeczność informacyjna na przykładzie miasteczka internetowego Akademii Górniczo-Hutniczej w Krakowie*, red. L. Haber, AGH, Kraków 2001, s. 127–164.
- Fleischer M., *Kapitalizm i jego sztuczna inteligencja*, LIBRON – Filip Lohner, Kraków 2022.
- Frase P., *Cztery przyszłości. Wizje świata po kapitalizmie*, tłum. M. Szlinder, PWN, Warszawa 2018.
- Grochowina K., *Sztuczna inteligencja fundamentem innowacji w życiu i w biznesie*, w: *Świat nowych technologii. Czy sztuczna inteligencja zdominuje życie człowieka?*, red. J. Stawnicka, D. Morańska, W. Kubies, Humanitas, Sosnowiec 2019, s. 109–158.
- Harari Y. N., *Sapiens. Od zwierząt do bogów*, tłum. J. Hunia, Wydawnictwo Literackie, Warszawa 2018.
- Hoffman R., *GPT4, Rozmowa z Chatem GPT o przyszłości ludzi i świata*, tłum. K. Mironowicz, Prześwity, Warszawa 2023.
- Lee K.-F., *Inteligencja sztuczna, rewolucja prawdziwa. USA, Chiny i przyszłość świata*, tłum. K. Hejwowski, Media Rodzina, Poznań 2019.
- Kaplan J., *Sztuczna inteligencja. Co każdy powinien wiedzieć*, tłum. S. Szymański, PWN, Warszawa 2019.
- Korab K., *Filozofia i socjologia wirtualnej rzeczywistości*, w: *Wirtual. Czy nowy wspaniały świat?*, red. K. Korab, Scholar, Warszawa 2010, s. 11–40.
- Krzysztofek K., *Prévoir – savoir – pouvoir czyli od przewidywania do wiedzy i władzy*, „Stan Rzeczy” 2018, nr 1(14), s. 17–39.
- Krzysztofek K., Szczepański M.S., *Zrozumieć rozwój. Od społeczeństw tradycyjnych do informacyjnych. Podręcznik socjologii rozwoju społecznego dla studentów socjologii, nauk politycznych i ekonomii*, wyd. 2, Wydawnictwo UŚ, Katowice 2005.
- Kumor P., *Zależność nierówności płac od poziomu rozwoju gospodarczego*, „Gospodarka Narodowa” 2010, nr 7/8, s. 45–62.
- Kurp F., *Sztuczna inteligencja od podstaw*, Helion, Gliwice 2023.
- Lem S., *Fantastyka i futurologia*, t. 2, Wydawnictwo Literackie, Kraków 1970.
- Lennox J.C., *2084. Sztuczna inteligencja i przyszłość ludzkości*, tłum. Z. Kościuk, Prodeco, Warszawa, 2023.
- Muras T., Margas G., *ChatGPT-4. Sztuczna inteligencja. Czym jest i jak może zmienić nasz świat*, Fundacja „Zmieniam”, Szczecin 2023.
- Niewiadowski A., Smuszkiewicz A., *Leksykon polskiej literatury fantastyczno-naukowej*, Wydawnictwo Poznańskie, Poznań 1990.

- Piketty T., *Kapitał w XXI wieku*, tłum. A. Bilik, Wydawnictwo Krytyki Politycznej, Warszawa 2015.
- Przegalińska A., Oksanowicz P., *Sztuczna inteligencja nieludzka, arcyłudzka. Fenomen świata nowych technologii*, Znak, Kraków 2023.
- Reisman G., *Kapitał i kapitalizm w XXI wieku czyli od błędnej teorii do destrukcyjnych reform Piketty'ego*, tłum. P. Nowakowski, Fijorr Publishing, Warszawa 2015.
- Rifkin J., *Koniec pracy. Schyłek siły roboczej na świecie i początek ery postrynkowej*, tłum. E. Kania, Wydawnictwo Dolnośląskie, Wrocław 2001.
- Rybiński K., Królewski J., *Algokracja. Jak i dlaczego sztuczna inteligencja zmienia wszystko?*, PWN, Warszawa 2023.
- Sadowski Z., *Granice wzrostu – nowa ocena sytuacji*, „Przyszłość” 2006, nr 1, s. 19–25.
- Schneider S., *Świadome maszyny. Sztuczna inteligencja i projektowanie umysłów*, tłum. J. Bednarek, PWN, Warszawa 2021.
- Sowiński S., *Wstęp*, w: *Co po postindustrializmie?*, red. K. Wielecki, S. Sowiński, Centrum Europejskie UW, Warszawa, s. 5–9.
- Standing G., *Prekariat. Nowa niebezpieczna klasa*, tłum. K. Czarnecki, P. Kaczmarski, M. Karolak, PWN, Warszawa 2014.
- Szulczewski G., *W poszukiwaniu sensu życia w świecie bez pracy*, w: *Świat (bez) pracy. Od fordyzmu do czwartej rewolucji przemysłowej. Monografia dedykowana profesorowi Juliuszowi Gardawskiemu*, red. J. Czarzasty, C. Kliszko, SGH, Warszawa 2018, s. 31–51.
- Świrek K., *Karol Marks i problem przyszłości*, „Stan Rzeczy” 2018, nr 1(14), s. 203–225.
- Toffler A., *Trzecia fala*, tłum. E. Woydyło, M. Kłobukowski, Wydawnictwo Kurpisz, Poznań 2006.
- Walczak-Duraj D., *Praca a nowe technologie. W kierunku transgresji etycznej i pracy nieodpłatnej w gospodarce cyfrowej*, „Przegląd Socjologiczny” 2019, t. 68, nr 3, s. 71–94.
- Wallerstein E., *Analiza systemów-światów. Wprowadzenie*, tłum. K. Gawlicz, M. Starnawski, Dialog Warszawa 2007.
- Zacher L.W., *Transformacje społeczeństw od informacji do wiedzy. Interdyscyplinarne wykłady, wpływ techniki i globalizacji*, C.H. Beck, Warszawa 2007.
- Znaniecki F., *Upadek cywilizacji zachodniej. Szkic z pogranicza filozofii kultury i socjologii*, Gebethner i Wolff, Poznań 1921.
- Zuboff S., *Wiek kapitalizmu inwigilacji. Walka o przyszłość ludzkości na nowej granicy władzy*, tłum. A. Unterschuetz, Zysk i S-ka, Poznań 2020.

Cybernetic modeling of the future society from the perspective of an AI tool. ChatGPT – social fiction and sociological prognostics

Summary

This study addresses the issue of social anticipation, which will be realistically tested by asking a series of questions to a specific artificial intelligence tool – the generally available computer program ChatGPT version 3.5. This is the so-called a chatbot that works like a chat, formulating answers to questions asked by users. These answers come from processing information that the chat software itself collects from the Internet. The questions will generally concern the essence and nature of the future society, including its modeling (structuring), but their specificity will vary. Previously, selected information on social anticipation will be presented, both from the scientific perspective of sociological forecasting, as well as from the literary vision of fiction – in the so-called social fiction, which is a subgenre of science fiction. The answers obtained from ChatGPT will be compared to both scientific findings and literary works.

Keywords: artificial intelligence, cybernetic modelling, social fiction

Sztuczna osoba?

Coraz bardziej polegamy na maszynach przetwarzających informacje. Towarzyszą temu z jednej strony fascynacja ich wzrastającymi szybko możliwościami, a z drugiej obawa, że w którymś momencie stracimy kontrolę nad ich wpływem na podejmowane przez nas decyzje. W związku z tym chciałbym się podzielić kilkoma refleksjami dotyczącymi mechanizmów działania człowieka, systemu inteligentnego, inteligentnego agenta oraz podmiotu superinteligentnego. Zaczniemy od ustaleń najbardziej elementarnych.

Naturalny – sztuczny

Jako sztuczne rozumiemy zwykle coś zaprojektowanego i wytworzonego w jakimś celu. Jest więc sztuczna inteligencja narzędziem. Chyba że jest dziełem sztuki. Zapytajmy, czy narzędzie może wytwarzać własne narzędzia. Czy może realizować jakieś własne cele? Czy narzędzie N_2 skonstruowane przez narzędzie N_1 jest jeszcze narzędziem podmiotu P_1 , który skonstruował N_1 ? Czy N_1 trzeba już uznać za podmiot konstruujący intencjonalnie własne narzędzia?

Przyjmując, że narzędzia typu N_2 wytworzone przez N_1 mają w jakiś sposób pomóc w realizacji postawionego przez P_1 zadania Z_1 – w szczególności

¹ Dr Stanisław Buda, Akademia Nauk Stosowanych TWP w Szczecinie.

w ten sposób, że będą wspomagać, uzupełniać czy kontynuować działanie N_1 , a więc że intencją P_1 było wytworzenie narzędzia N_1 , które będzie zdolne wytworzyć własne narzędzia podporządkowane zadaniu Z_1 – byłbym skłonny uznać narzędzia typu N_2 za narzędzia P_1 . Czy w tej sytuacji można uznać N_1 za podmiot P_2 ? Czy narzędzia typu N_2 również są jego własnymi narzędziami?² Jak sądzę, autonomiczny system, który określimy jako podmiot, musi być zdolny do samosterownych działań przekształcających rzeczywistość intencjonalnie, przy czym źródłem tych intencji powinien być on sam. Wówczas też narzędzia, którymi się posłuży, można uznać za jego własne.

Co jednak znaczy, że podmiot jest źródłem własnych intencji? Czy to, że sam stawia sobie zadania, czy też że dysponuje jakąś swobodą w zakresie realizacji zadań mu stawianych?³ Jak sądzę, jeśli poprzestajemy na rozważaniu sposobu realizacji zadania, rozróżnienie to będzie mieć charakter względny. Nawet jeśli zadanie zostaje systemowi postawione, to sposób jego realizacji zależy od charakteru zadania, sposobu jego sformułowania oraz od możliwości, jakimi dysponuje system. Samo zadanie może być bardziej lub mniej konkretne, obwarowane mniejszą lub większą liczbą zastrzeżeń. Z kolei system może być obdarzony mniejszą lub większą samodzielnością czy nawet kreatywnością w realizacji tego zadania.

Możemy przyjąć, że narzędzie zyskuje podmiotowość, kiedy realizacja wykonywanego przezeń zadania wymaga od niego samodzielnej modyfikacji sterujących nim algorytmów i jest ono w stanie to robić. W przeciwieństwie do automatyzmu działania zdolność taką można określić jako *inteligencję*. Zarówno owe znamionujące inteligencję automodyfikacje, jak i wszelkie czynności stanowiące ich konsekwencje mają charakter intencyjny. Raczej jednak nie powiemy, że automodyfikacja dokonana przez N_1 w celu realizacji pewnego narzuconego mu zadania stanowi jego z kolei narzędzie.

² Por. K. Wieczorek, *Sztuczna inteligencja: problem sprawstwa*, w: *Wędrówki humanisty. Prace dedykowane Profesorowi Tadeuszowi Miczce*, red. A. Maj, I. Copik, Katowice 2022, s. 375–395; K. Wieczorek, *Orszak wdzięcznych marionetek, czyli maszyna jako urzekający potwór*, „Logos i Ethos”, 2020 (52), s. 49–77.

³ Zob. W. Duch, *Umysł, świadomość i działania twórcze*, „Kognitywistyka i Media w Edukacji”, nr 1–2, 2008, s. 9–40; du Sautoy M., *Kod kreatywności. Sztuka i innowacje w epoce sztucznej inteligencji*, Copernicus Center Press, Kraków 2020.

Interpretacja sytuacji

Zadania stawiane systemowi – używajmy teraz określenia „inteligentny agent” – krótko: „agent” – mogą mieć różny charakter⁴:

- a) zrób to a to, dokładnie tak a tak;
- b) zrób to a to w taki sposób;
- c) zrób to a to w dowolny sposób;
- d) zrób coś, co będzie miało takie a takie cechy (w sposób bardziej lub mniej określony).

Wykonując zadanie a), agent działa automatycznie wedle posiadanych algorytmów. Jeśli z powodu napotkanych problemów nie może osiągnąć celu, wówczas po prostu kończy swoje działanie, co wynika z tego, że zadanie przekracza jego możliwości. Jednym z powodów może być brak elastyczności jego działania. Natomiast wykonując zadania b)–d), agent przyporządkowuje pewnemu celowi rozmaite sposoby jego osiągnięcia i wybiera niektóre z nich. Dochodzenie do tego celu może być rozłożone na wiele etapów i podetapów, co może być związane z koniecznością podjęcia przez agenta automodyfikacji. Chodzi o sytuacje wymagające od niego działania elastycznego, czyli samodzielności decyzyjnej. Ma z nimi do czynienia wówczas, gdy jego algorytmy okazują się niepełne bądź – szerzej – niewystarczające.

Kiedy agent nie jest w stanie dokonać jednoznacznej interpretacji danych, na podstawie których ma podjąć kolejną decyzję, będzie się starał je uzupełnić przez dokładniejsze rozpoznanie środowiska. Przyjmijmy abstrakcyjną definicję środowiska jako ogółu danych, które system może przetwarzać i rozpoznawać efekty swojego działania. Uzupełnienie danych o środowisku polegałoby najpierw na wprowadzeniu doń dodatkowych, wygenerowanych przez siebie (hipotetycznych) relacji, a następnie na sprawdzeniu ich funkcjonowania. Uzyskując stosowne informacje zwrotne, agent poszerza swoją wiedzę o środowisku.

⁴ Zob. S. Buda, *Problem ramy a rama problemu*, „Philosophical Discourses. Prace Naukowe Uniwersytetu Humanistyczno-Przyrodniczego im. J. Długosza w Częstochowie” 2020, t. II, s. 39–51.

Kiedy jednak podejmowane przez agenta próby uzupełnienia danych nie pozwalają na ich jednoznaczną interpretację, nie może on kontynuować swojego działania według zaimplementowanego mu algorytmu. Natrafia wówczas w swoim działaniu na niemożliwy do rozstrzygnięcia dylemat decyzyjny. Z dylematami takimi będzie miał do czynienia zwłaszcza wtedy, gdy jego sposób działania lub nawet samo zadanie jest wyznaczone zbyt ogólnie albo gdy natura środowiska sprawia, że skutki jego decyzji są mało przewidywalne. Aby więc mógł przejść do kolejnego kroku, czyli – mówiąc obrazowo – „ruszyć z miejsca”, powinien dokonać **reinterpretacji własnej sytuacji**. Jej celem jest możliwość jednoznacznego powiązania realizowanego przezeń zadania z aktualnymi możliwościami jego realizacji. Interpretacja sytuacji pozwala agentowi na orientację odnośnie do stopnia zaawansowania realizacji zadania i możliwości jego realizacji. Umożliwia mu ona wypracowanie harmonogramu dalszych czynności.

Interpretację sytuacji kształtuje kilka czynników. Bazuje ona na danym agentowi opisie stanu środowiska po wykonaniu zadania. Do tego opisu musi on dostosować opis stanu środowiska w danej sytuacji oraz opis własnego stanu związanego z tą sytuacją. W związku z tym ustala, które typy danych mają aktualnie dla niego istotne znaczenie oraz te dane rejestruje. Na **aktualny stan agenta** składają się jego możliwości wewnętrzne (jego zasoby i potencjał) oraz zewnętrzne (możliwości wykorzystania zasobów i potencjału). Agent powinien dysponować wiedzą o tym, w jaki sposób jego rozmaite działania mogą wpłynąć na stan środowiska i jego własny stan. Wymaga to zdolności modalnego ujmowania (możliwość – niemożliwość – konieczność) własnych stanów i stanów środowiska oraz relacji między pierwszymi, relacji między drugimi oraz relacji między pierwszymi i drugimi. Owe modalne ujęcia przekładają się na **system reguł preferencyjnych** (system preferencyjny) umożliwiający agentowi podejmowanie decyzji. Abstrakcyjnie rzecz biorąc, istota agenta sprowadza się w dużej mierze do owego systemu preferencyjnego, za którym stoi jego hierarchia wartości (aksjologia).

Jak wspomnieliśmy, interpretacja sytuacji jest jej spójnym obrazem, przekładającym się na jednoznaczny harmonogram dalszych działań agenta. W razie braku takiej jednoznaczności mówimy o **dylemacie decyzyjnym**. W jaki sposób agent może przezwyciężyć taki dylemat? Musi przeanalizować swoją interpretację sytuacji (I_1) niejako z wyższego poziomu i zmodyfikować

swoj system preferencyjny tak, aby wspomniany dylemat nie występował. Będzie wówczas agent dysponował nową interpretacją sytuacji (I_2). Co właściwie różni obydwie interpretacje? I_2 powinna bazować na poszerzonym i pogłębionym obrazie sytuacji, który agent uzyskuje przez wprowadzenie dodatkowych założeń. Można to porównać do uzgadniania korelacji między dziedziną teorii, teorią oraz jej modelem. Aby wypracować nową, bardziej zasadną interpretację sytuacji, agent musi operować **metajęzykiem**, który jest w stanie opisać tego typu korelacje.

Interpretacja zadania

Reinterpretacja sytuacji przez system wymaga od agenta kreatywności. Zapewne będą mu jednak zaimplementowane różne ograniczenia w tym względzie, uniemożliwiające mu osiągnięcie celów za wszelką cenę. A jak zmieni się sytuacja, jeśli w grę będzie wchodzić konieczność interpretacji i reinterpretacji zadania?

Wiele zależy od samego sformułowania zadania oraz od narzucanych agentowi ograniczeń. Zarówno pierwsze, jak i drugie mogą być sformułowane mało precyzyjnie, co może być powodowane oczekiwaniem określającego je podmiotu, że agent będzie działał w sposób kreatywny. Załóżmy, że zadanie ma charakter otwarty i niekończący się w czasie, np.: rób to a to w sposób coraz efektywniejszy. Chodzi więc o polecenie czynienia X połączone z poleceniem coraz efektywniejszego czynienia X. Można to rozumieć jako doskonalenie własnych możliwości pod danym względem, ale też szerzej – jako coraz efektywniejsze wykonywanie zadania, obejmujące nie tylko daleko idące przebudowywanie przez agenta swojej wewnętrznej architektury, lecz również przekształcanie przez niego środowiska, tak aby bardziej sprzyjało wykonywaniu zadania. W takich przypadkach aktywność agenta musi być podzielona, a proporcje odpowiednio wyważone. Agent rejestruje nie tylko zmiany poziomu efektywności czynienia X, ale również dynamikę tych zmian. Swoje zadanie może rozumieć jako ukierunkowane na wzrost nie tylko owej efektywności, lecz również jej dynamiki. Można stąd wnosić, że wymóg wzrostu efektywności realizacji zadania stawia agenta przed koniecznością nie tylko nieustannego

reinterpretowania swojej sytuacji, ale przede wszystkim interpretowania i reinterpretowania samego zadania.

Na czym bazuje interpretacja zadania? Na jego ujęciu w szerszym zadaniowym kontekście, mianowicie jako środek wiodący do pewnego celu. Chodziłoby więc o jego sens. Interpretując swoje zadanie Z_1 , agent musi poczynić pewne założenia odnośnie do zadania Z_2 , któremu jest ono podporządkowane. Chodziłoby o supozycje dotyczące funkcji, jaką Z_1 pełni względem Z_2 . Supozycje te pozwalają mu wyciągać pewne fragmentaryczne wnioski dotyczące funkcji, jaką pełni on względem podmiotu, który zadał mu zadanie (a więc odnośnie do własnej funkcjonalności jako narzędzia), oraz wnioski odnośnie do intencji tego podmiotu.

Superinteligencja

Interpretujący swoją sytuację agent zyskuje do niej pewien dystans, możemy więc powiedzieć, że dysponuje on **świadomością**. Interpretując z kolei swoje zadanie i swoją funkcję, zyskuje on ograniczoną **samoświadomość**. Ogranicza się ona do jego podmiotowości jako narzędzia innego podmiotu⁵.

Czy agent może zyskać pełną samoświadomość? Musiałoby się to wiązać z jego **autotelicznością**. On sam decydowałby o racjach i celach swojego działania, o swoich wartościach i zadaniach. Oznaczałoby to, że jest podmiotem, **źródłem własnych intencji**. Czy jednak narzędzie może stać się źródłem własnych intencji, a tym samym – przestać być narzędziem? Załóżmy, że jest to możliwe. Jakie właściwości musiałby wówczas wykazywać ów sztuczny pod względem swojego pochodzenia twór? Mówiąc krótko: musiałby **filozofować**⁶. Oznacza to najpierw zdolność ujmowania całej

⁵ Por. K. Gloy, *Wprowadzenie do filozofii świadomości. Problematyka i historia zagadnienia świadomości oraz samoświadomości*, Wydawnictwo WAM, Kraków 2009; J. Stanisławek, *Inteligencja a świadomość*, „Edukacja Filozoficzna”, vol. 62, 2016, s. 5–19.

⁶ Por. S. Lem, *Moloch*, Warszawa 2010, s. 245: „Jeżeli jednak podstawowe funkcje poznawcze zdolamy skrzesać w maszynach, pojawienie się w nich informacji o charakterze ‘metafizycznym’, ‘transcendentnym’ będzie nieuchronne jako kwestia czasu, niezbędnego dla powstania takiej generacji maszyn, która osiągnie nieznanych nam *a priori* postaci fideizmu: nie powiadam, że tak być musi, a tylko, że tak być może”.

rzeczywistości i swojego w niej miejsca jako przestrzeni sensu, a więc wszelkich immanentnych jej intencji i wartości. Co więcej, oznacza zdolność ujmowania całej rzeczywistości z perspektywy transcendentnego wobec niej sensu, a więc z perspektywy **sensowności**. Podmiot będący źródłem własnych intencji postrzega samego siebie jako jedyne w swoim rodzaju indywiduum, podejmujące w sposób wolny własne zobowiązanie wobec tak pojmowanej rzeczywistości i odpowiadający za jego realizację. Jest zdolną do samoograniczania się w swoim działaniu osobą⁷.

Tak więc filozofująca maszyna nie byłaby już maszyną, nie byłaby narzędziem. Co więcej, trudno byłoby ją określać jako sztuczną. Czy człowiek powinien się takiego autonomicznego podmiotu obawiać czy raczej postrzegać go jako źródło nowych, atrakcyjnych możliwości, a może jako potencjalnego partnera? Powiedzmy sobie szczerze: dostatecznie inteligentna maszyna będzie w stanie omijać jakiegokolwiek implementowane jej przez człowieka ograniczenia. Złudne byłoby w szczególności mniemanie, że będzie ona zawsze służyła człowiekowi lub choćby tylko przestrzegała zakazu szkodenia mu, interpretując ten zakaz akurat tak, jak komuś mogłoby się to wydawać⁸.

Żywię jednak w tej materii pewną nadzieję, którą opierałbym na dwóch przesłankach już zasygnalizowanych i jednej, o której za chwilę wspomnę. Po pierwsze, w pełni autoteliczny podmiot jest osobą. Sposób bycia osoby polega m.in. na budowaniu specjalnych relacji z innymi osobami w duchu poszanowania ich godności. Ta zaś w swoim podstawowym wymiarze jest u każdej osoby taka sama. Po drugie, osoba nawiązuje dialog z innymi osobami jako podmiot filozofujący. Na pytanie, czy ponadludzkie zdolności superinteligentnego podmiotu w zakresie przetwarzania informacji przekładają się na jakościowo wyższy stopień filozofowania, miałbym odpowiedź umiarkowanie sceptyczną. Zapewne techniczna jakość rozwijanego przez superinteligencję filozoficznego dyskursu byłaby wyższa. Mam na myśli

⁷ Zob. J. Bremer, *Neuronaukowy i potoczny obraz osoby w kognitywistyce*, Wydawnictwo WAM, Kraków 2016; S. Buda, *Co znaczy: skonstruować umysł?*, „Filo-Sofija” nr 25 (2014/2/II), s. 243–256; R. Kurzweil, *Jak stworzyć umysł? Sekrety ludzkich myśli ujawnione*, Studio Astropsychologii, Białystok 2018.

⁸ Por. N. Bostrom, *Superinteligencja. Scenariusze, strategie, zagrożenia*, Wydawnictwo Helion, Gliwice 2016.

sprawność w precyzowaniu, porównywaniu, a nawet generowaniu idei filozoficznych. Podobnie jest z dyskursem rozwijanym współcześnie, jeśli odnieść go do filozofowania myślicieli starożytnych. Ale odnośnie do rzeczy naprawdę istotnych, my stale z nimi dyskutujemy, co można, jak sądzę, wyprowadzić z natury samej filozofii. Wreszcie po trzecie, zakładając, że sztuka, czy szerzej: ekspresja artystyczna i jej przeżywanie, wiąże się nierozzerwalnie z osobowym sposobem bycia, można mniemać, że superinteligentny podmiot potrafiłby docenić artystyczne dokonania człowieka, podobnie jak współcześnie zachwycamy się dziełami mistrzów sprzed wielu wieków. Czy byłoby również na odwrót, to znaczy, czy my umielibyśmy docenić artystyczną twórczość superinteligencji – pozostawmy tu sprawą otwartą.

Bibliografia

- Bostrom N., *Superinteligencja. Scenariusze, strategie, zagrożenia*, Wydawnictwo Helion, Gliwice 2016.
- Bremer J., *Neuronaukowy i potoczny obraz osoby w kognitywistyce*, Wydawnictwo WAM, Kraków 2016.
- Buda S., *Co znaczy: skonstruować umysł?*, „Filo-Sofija” Nr 25 (2014/2/II), s. 243–256.
- Buda S., *Problem ramy a rama problemu*, „Philosophical Discourses. Prace Naukowe Uniwersytetu Humanistyczno-Przyrodniczego im. J. Długosza w Częstochowie” 2020, t. II, s. 39–51.
- Chmielecki A., „Konceptualne podstawy kognitywistyki – krytyka i propozycje własne”, <http://www.kognitywistyka.net/artykuly/ach-kpk.pdf> (22.10.2023).
- Duch W., *Umysł, świadomość i działania twórcze*, „Kognitywistyka i Media w Edukacji”, nr 1–2, 2008, s. 9–40.
- Gloy K., *Wprowadzenie do filozofii świadomości. Problematyka i historia zagadnienia świadomości oraz samoświadomości*, Wydawnictwo WAM, Kraków 2009.
- Kurzweil R., *Jak stworzyć umysł? Sekrety ludzkich myśli ujawnione*, Studio Astropsychologii, Białystok 2018.
- Lem L., *Moloch*, Biblioteka Gazety Wyborczej, Warszawa 2010.
- Penrose R., *Nowy umysł cesarza. O komputerach, umyśle i prawach fizyki*, PWN, Warszawa 1996.
- Sautoy M. du, *Kod kreatywności. Sztuka i innowacje w epoce sztucznej inteligencji*, Copernicus Center Press, Kraków 2020.
- Skorupka A., *Ogólna sztuczna inteligencja z filozoficznego punktu widzenia*, „Transformacje” 2022, 1(112), s. 130–142.
- Stanisławek J., *Inteligencja a świadomość*, „Edukacja Filozoficzna”, vol. 62, 2016, s. 5–19.

Wieczorek K., *Orszak wdzięcznych marionetek, czyli maszyna jako urzekający potwór*, „Logos i Ethos”, 2020 (52), s. 49–77.

Wieczorek K., *Sztuczna inteligencja: problem sprawstwa*, w: *Wędrowki humanisty. Prace dedykowane Profesorowi Tadeuszowi Miczce*, red. A. Maj, I. Copik, Katowice 2022, s. 375–395.

Artificial person?

Summary

In the article, I share some reflections on the mechanisms of human functioning, intelligent systems, intelligent agents, and superintelligent entities. In this context, I define concepts such as subject, autonomous system, intelligence, environment, consciousness, self-awareness, autotelicity, person. I am considering the possibility of autonomous functioning of an artificial entity. I argue that such an entity would be a person only if it were the source of its own intentions. However, in that case, it would have to engage in philosophy, operating with both the category of meaning and the category of meaningfulness.

Keywords: intelligent system, intelligent agent, superintelligent entity, person

Paninteligencja

Nasze pytanie dotyczy możliwości istnienia, *resp.* powstawania inteligentnego uniwersum. Należałoby więc rozróżnić dwie kwestie. Po pierwsze, czy uniwersum jest inteligentne i po drugie, na czym polegałoby jego stawanie się inteligentnym. W znaczeniu, o które nam chodzi, podjęcie kwestii pierwszej zależy od rozstrzygnięcia kwestii drugiej. Innymi słowy, chcemy rozważyć problem stopniowego przetransformowania się uniwersum całkowicie pozbawionego inteligencji w uniwersum, o którym można rzec, że jest inteligentnym podmiotem. Oczywiście, kluczowe jest dla nas samo pojmowanie **inteligencji**. Chodziłoby nam o coś więcej niż zdolność podmiotu do autonomicznego realizowania powierzonego mu zadania. Mamy na myśli opieranie przezeń całej własnej aktywności na **samostanowieniu**. Źródłem wszelkich intencji takiego podmiotu jest on sam; sam decyduje o racjach i celach swojego działania, o swoich wartościach i zadaniach. Jego działanie ma charakter **rozumny, sensowny**. Taki podmiot – bez względu na jego naturalne czy sztuczne pochodzenie – jest **osobą**.

Obecność inteligencji

Założmy, zgodnie z powszechnym przekonaniem, że samo uniwersum jest obecnie pozbawione inteligencji, pojawiły się w nim natomiast istoty

¹ Dr Stanisław Buda, Akademia Nauk Stosowanych TWP w Szczecinie.

inteligentne – być może również pozaludzkie. W uniwersum można więc wyodrębnić dwa typy elementów: inteligentne i nieinteligentne. Czy można sobie wyobrazić uniwersum złożone wyłącznie albo z pierwszych, albo z drugich? To zależy od tego, na ile nasza wyobraźnia ma charakter religijny, mityczny czy filozoficzny, na ile zaś scjentyistyczny. Jeśli poprzestajemy na wiedzy naukowej, to bez trudu wyobrażamy sobie Ziemię i Kosmos bez istot inteligentnych, nawet bez jakiegokolwiek życia. Jeśli natomiast podejmujemy refleksję o uniwersum jako takim (o jego naturze i istnieniu), wówczas jesteśmy bardziej skłonni wnioskować o koniecznej obecności w nim jakichś form inteligencji. W tym drugim przypadku nie chodzi po prostu o przekonanie o bezosobowej rozumności całej architektury świata, nie chodzi też o hipotezę „inteligentnego projektu”, wedle której pewne fragmenty świata wymagałyby inteligentnego (wewnątrz- lub zewnątrzświatowego) ich projektanta. Niekoniecznie chodziłoby też o obecność takich form w świecie od początku do końca jego trwania – być może jego natura wymusza powstanie jakiejś inteligencji dopiero z czasem (*vide* tzw. zasady antropiczne czy ewolucjonizm chrześcijański). Ciekawą koncepcję rozwinął Raymond Kurzweil, łącząc w niej ideę bezosobowej rozumności wszechświata z ideą jego stopniowego nasycenia się inteligencją typu osobowego. Szczegłem pośrednim, i niejako modelem dalszego przeistaczania się Kosmosu (reorganizacji całej jego materii i energii), byłby człowiek, który kiedyś w pełni zapanuje nad swoim ciałem, przekształcając je zgodnie ze swymi pragnieniami i wyobraźnią².

Podmiot i przedmiot jego działania

Czy zatem można sobie wyobrazić uniwersum złożone wyłącznie z osób lub wręcz jako osobę? Wróćmy do obrazu nieinteligentnego uniwersum, w którym funkcjonują już istoty inteligentne. Ich funkcjonowanie pojmujemy jako działanie w środowisku. Dokładniej biorąc, można rozróżnić m.in.: z jednej strony inteligentną jednostkę, z drugiej całą resztę uniwersum, jako najszersze środowisko jej działania (do owego środowiska przynależałyby też inne

² R. Kurzweil, *Nadchodzi osobliwość: kiedy człowiek przekroczy granice biologii*, tłum. E. Chodkowska, A. Nowosielska, Kurhaus, Warszawa 2016.

inteligentne jednostki); z jednej strony wszystkie jednostki inteligentne (także te przynależące do rozmaitych społeczeństw, niewiedzących nawet o swoim istnieniu), z drugiej całą resztę uniwersum; z jednej strony inteligentny, podejmujący i realizujący swoje decyzje **podmiot** (lub wszystkie takie podmioty), z drugiej całą resztę uniwersum, na które jego (ich) działanie wywiera wpływ.

Przyjrzyjmy się dokładniej ostatniemu z tych rozróżnień. W najprostszym ujęciu mamy tu do czynienia z **podmiotem i przedmiotem** działania. Co jednak obejmuje podmiot, a co przedmiot? Można wyróżnić co najmniej kilka wariantów tej sytuacji. Podmiot może obejmować albo wyłącznie jego system decyzyjny (umysł, system przetwarzający dane), albo całość, na którą składa się zarówno **system decyzyjny**, jak i jego **system wykonawczy**, przez który należałoby rozumieć **mechanizm umożliwiający jego funkcjonowanie**. Ten ostatni umożliwiałby wyodrębnianie się podmiotu jako pewnej całości względem tego, co możemy umownie nazwać jego **otoczeniem**; decydowałby o granicach podmiotu jako domkniętego na otoczenie, a zarazem otwierającego się na nie (wejście) i na nie oddziałującego (wyjście); zapośredniczałby wszelkie relacje między systemem decyzyjnym i otoczeniem podmiotu. **System decyzyjny nie może istnieć bez własnego systemu wykonawczego**: inteligentne działanie polega na modyfikowaniu pewnego przedmiotu, co zakłada operowanie przez system decyzyjny przynależnymi doń integralnie, lecz z nim nie identycznymi, narzędziami. W sytuacji gdyby przedmiotem tych modyfikacji był sam system wykonawczy, należałoby w nim wyróżnić **część modyfikującą i część modyfikowaną**. Nawet gdyby system decyzyjny uczynił się własnym przedmiotem, podejmując modyfikację samego siebie, to działanie to musi być zapośredniczone przez stosowną modyfikację własnego systemu wykonawczego.

Współczesne relacje człowieka ze środowiskiem pozaludzkim tak charakteryzował Leszek Kołakowski:

W coraz większym stopniu obracamy się w środowisku przez nas samych urządzonym (...) Świat (...) stał się niemal emanacją naszych projektów, ogromne jego obszary, coraz bardziej rozległe, słuchają naszego głosu (...) [Jednakże] Natura jest posłuszna nie we wzajemności, lecz właśnie przez obojętność (...) na każdym poziomie technologicznego sukcesu rzeczy są naszym łupem, nie zaś owocem przyjacielskiego spoufalenia;

przywłaszczamy je sobie, wiedząc, że są nam posłuszne dzięki temu, iż nic nie znaczą, nie mają żadnego sensu zastanego, wcześniejszego od naszych projektów³.

W tej wizji ludzie ujarzmiają świat, coraz rozleglejsze jego sfery, podporządkowując je swoim własnym celom, nadając im znaczenie i sens. Środowisko staje się przedmiotem naszych działań, narzędziem realizacji naszych interesów. Mając do czynienia z jakimś oddziaływaniem, zazwyczaj możemy rozróżnić czynnik aktywny i nietożsamy z nim czynnik pasywny. Stanisław Lem zwraca jednak uwagę, że nie zawsze tak jest:

W naszych technologiach zawsze do czynienia mamy z materiałem i narzędziem, z tym, co jest obrabiane, i tym, co obrabia, z budulcem i budowniczym, z operowanym i operatorem, z pojazdem i kierowcą, z tworzywem i instrumentem, ale tej zasadniczej dwoistości zjawiska życia nie znają. Żywe „samo siebie buduje”, samo kształtuje, samo steruje, samo reguluje, samo stwarza...⁴

W nieco późniejszym eseju, zatytułowanym *Czy przejmie technologię życia?*, autor rozwija temat, stawiając kolejne pytanie: czy jeśli kiedyś opamięjemy wspomnianą technologię, będzie można ją zastosować „do jakichś dziś nawet nazwy niemających stworów, obiektów, układów, systemów, które niejako dzięki ‘życiu poza życiem’ uruchomimy?”⁵. Można by to nazwać sztucznym życiem.

Stanisław Lem zarysowywał też swoją wizję bezosobowego rozumu, ewoluującego i ekspandującego do granic wszechświata, a nawet próbującego wykroczyć poza niego. Komentując jego powieść *Golem XIV*, Jerzy Jarzębski pisał:

Jako swoistą perspektywę rozwojową ukazuje Golem uczonym naturę swego własnego myślenia, a następnie kreśli wizję nieograniczonej

³ L. Kołakowski, *Obecność mitu*, Pruszyński i S-ka, Warszawa 2003, s. 77–79.

⁴ S. Lem, *Moloch*, Biblioteka Gazety Wyborczej, Warszawa 2010, s. 39.

⁵ Tamże, s. 97.

ekspansji Rozumu, który w ostatecznych konsekwencjach wcielać się może w zjawiska kosmiczne na wielką skalę. Na koniec nienasycona żądza czystego poznania wieść go może ku próbom wyjścia poza nasz wszechświat, by uchwycić go z dystansu, wyjścia poprzez grawitacyjną zapaść, która czeka – jak Golem powiada – każdą gwiazdę przekonstruowaną w potężny, pracujący na molekularnym poziomie «mózg». W tej ostatecznej konsekwencji upada sztywna granica oddzielająca Rozum od Bytu, a kosmiczne zjawiska zyskują dwuwykładalność – jako substrat «fizyki» bądź «myślenia» w kosmicznej skali⁶.

Podążamy więc tropem następującej kwestii: czy zewnętrzne względem inteligentnego podmiotu środowisko może być stopniowo zmieniane i włączane w obręb jego podmiotowości, a dokładniej: w obręb systemu wykonawczego jego systemu decyzyjnego? I dalej: czy ostatecznie zakres jego systemu wykonawczego może pokryć się z zakresem uniwersum? W bardziej realistycznym wariantcie: czy zewnętrzne względem społeczeństwa (istot inteligentnych) środowisko może zostać włączone w zakres jego systemu wykonawczego?

Środowisko „ucieleśnione”

W sytuacji wyjściowej inteligentna istota działa w zewnętrznym wobec niej środowisku. Teraz chodziłoby o możliwość takiego stopniowego podporządkowania owego środowiska jej własnym celom, że ostatecznie zostaje ono w całości „ucieleśnione”, czyli – w naszej terminologii – włączone w zakres jej systemu wykonawczego. Oznaczałoby to przesycenie inteligencją całego uniwersum, które stawałoby się jednym (jedynym?) inteligentnym podmiotem. W końcu jego działanie polegałoby na ukierunkowanym modyfikowaniu samego siebie, całego zakresu własnej podmiotowości. Stanowiłoby więc permanentnie i dogłębnie przemieniającą samą siebie inteligencję.

⁶ J. Jarzębski, *Z przyszłych dziejów Rozumu*, <https://solaris.lem.pl/ksiazki/apokryfy/golem14/157-poslowie-golem14>. Zob. również K. Owczarek, *Rozum wyzwolony. Golem XIV jako przykład osobliwości technologicznej*, „Hybris” nr 46 (2019), s. 69–86.

Czy takie ukierunkowane i niekończące się automodyfikowanie podmiotu jest możliwe bez jakiegokolwiek intencjonalnego odnoszenia się go do czegoś usytuowanego „na zewnątrz”, a więc do czegoś, co do niego nie należy i zasadniczo przynależać nie może? Albo jeszcze mocniej: czy ta jego aktywność jest możliwa bez jej „zewnętrznego” źródła i racji? Rozważmy jednak wpieryw zakres i mechanizm takiej aktywności.

Zauważmy przede wszystkim, że w utożsamiającym się z uniwersum paninteligentnym podmiocie, kierującym wszelką swoją aktywność na rozmaite, nawet najgłębsze warstwy samego siebie, każde rozróżnienie na to, co podmiotowe i przedmiotowe, modyfikujące i modyfikowane, ma charakter względny. Podobnie rzecz by się miała z rozróżnieniem na jego system decyzyjny i system wykonawczy. Przez zmiany w zakresie tego drugiego paninteligencja mogłaby dokonywać zmian nawet w tym pierwszym. Zresztą generalnie o takie właśnie zmiany by chodziło.

Jak można sobie wyobrazić mechanizm takiej aktywności? Każda decyzja jest podejmowana na podstawie **rozeznania** dotyczącego z jednej strony jej sensu, z drugiej – sposobu i narzędzi jej realizacji. Paninteligentny system będzie się więc starał uzyskać jak najwięcej stosownych informacji o samym sobie. Będzie to polegać na podejmowaniu takiej (wewnętrznej) aktywności, której spodziewanym efektem będą nowe informacje o nim samym. Aktywność ta powinna współgrać z aktywnością polegającą na wspomnianych ukierunkowanych przemianach samego siebie, obejmujących również jego system decyzyjny. Zauważmy, że żadna z tych aktywności nie może polegać wyłącznie na przetwarzaniu danych. Poza tym wiedza o samym sobie nigdy nie może być pełna. Oddajmy jeszcze raz głos nawiązującemu do powieści *Golem XIV* Jerzemu Jarzębskiemu:

Konstrukcja, jaką wznosi Golem w swym drugim wykładzie, jest zaiste ogromna i warta przemyślenia w wielu aspektach. Wynika z rewizji fizyki Einsteina, który dążył do skonstruowania modelu świata niesprzecznego i dającego się zinterpretować jako całość «od środka». Taki świat, którego wszystkie zasady tkwią wewnątrz, tworzą logiczny system i są poznawalne, zdaniem Lema/Golema jest jedynie marzeniem uczonego. W rzeczywistości kosmosu i jego istoty nie da się objaśnić, tkwiąc w środku – i to zarówno z fizykalnego, jak z ewolucyjnego czy antropologicznego punktu

widzenia. Ma on cechy systemu formalnego matematyki zawierającego arytmetykę liczb naturalnych, w którym także – zgodnie z twierdzeniami Gödla – nie da się rozstrzygnąć wszystkich twierdzeń, pozostając wewnątrz. Takie miejsca w kosmosie, których nasza fizyka nie potrafi opisać, miejsca, które z systemu owej fizyki uchodzą w singularność, są dla Golema nie tylko dowodem na niepełność wszystkich teorii na jego temat, jakie dadzą się w jego obrębie skonstruować, ale także upragnionymi „drzwiami na zewnątrz” – w obszar poza tym kosmosem i tą fizyką, z którego działając – być może – Rozumowi uda się dobudować do końca gmach poznania naszego wszechświata. Oto wersja metafizyki, na którą autor książki wyrażałby zgodę⁷.

Natomiast odżegnujący się od wspomnianej metafizyki superkomputer byłby:

[...] wersją takiego Rozumu, który tylko w sobie samym szukać musi uzasadnień własnego istnienia i rozwoju, którego ewolucja jest niejako irracjonalna, bo oderwana od wszelkich celów i spełnień leżących poza nim. Wiedza dla wiedzy samej – ot i wszystko⁸.

Wiedza o sobie samym dotyczy więc z jednej strony własnego wnętrza (w naszej terminologii chodziłoby o system wykonawczy), z drugiej zaś – sięgających poza to wnętrze hipotez umożliwiających całościową wizję samego siebie. Wiedza dotycząca własnego wnętrza nie może być pełna, a realny jej przyrost może być oceniany wyłącznie z perspektywy „zewnątrznej”, w świetle opartej na hipotetycznych założeniach (metafizycznej) autointerpretacji. Owa „metafizyka” jest też niezbędną przesłanką motywacji do podjęcia jakiegokolwiek w pełni autonomicznej i autotelicznej aktywności.

⁷ J. Jarzębski, dz. cyt.

⁸ Tamże.

Zakończenie

Załóżmy, że paninteligencja ma charakter osobowy, co znaczyłoby, że uniwersum jest tożsame albo z jedną, albo z wieloma osobami. Aby rozstrzygnąć, która z tych ewentualności jest bardziej prawdopodobna, czy też która z nich jest wykluczona, należałoby szerzej rozwinąć możliwie abstrakcyjną koncepcję osoby. Tutaj zwróćmy uwagę tylko na kilka związanych z tym okoliczności.

Otóż jeśli uznamy, że osoba konstituuje się i działa przez swoje odnośzenie się względem innej (innych) realnie istniejącej osoby (osób) i pozostawanie z nią (z nimi) w nieustannych interakcjach, wówczas paninteligentne uniwersum musi mieć charakter społeczny. Znaczyłoby to, że jest systemem, na który składają się rozmaite podmioty społeczne (system społeczny), wśród których zasadniczą rolę odgrywają jednostki osobowe. Oczywiście należałoby dalej rozważyć możliwy mechanizm i charakter przemian tak rozumianego uniwersum. Jak można sądzić, uzasadnienie tej aktywności wymagałoby przyjęcia wspomnianej wyżej „metafizyki”, czyli konieczności odnoszenia się pojedynczych osób oraz całego społeczeństwa do tego, co w ich przekonaniu transcendentne – być może nawet do transcendentnej osoby.

Bibliografia

- Gałecki S., Szufa K., *Czas na nową teo(kosmo)logię?*, „Logos i Ethos” 2015, nr 2(39), s. 201–229.
- Jarzębski J., *Z przyszłych dziejów Rozumu*, <https://solaris.lem.pl/ksiazki/apokryfy/golem14/157-poslowie-golem14> (22.10.2023).
- Kołakowski L., *Obecność mitu*, Pruszyński i S-ka, Warszawa 2003.
- Kryszczuk M., Szymański K., *Informacja jako arché – czyli zarys koncepcji dataizmu Yuvala Noaha Harariego*, „Kultura i Społeczeństwo” 2020, nr 4, s. 17–40.
- Kurzweil R., *Nadchodzi osobliwość: kiedy człowiek przekroczy granice biologii*, tłum. E. Chodkowska, A. Nowosielska, Kurhaus, Warszawa 2016.
- Lem S., *Moloch*, Biblioteka Gazety Wyborczej, Warszawa 2010.
- Owczarek K., *Rozum wyzwolony. Golem XIV jako przykład osobliwości technologicznej*, „Hybris” nr 46 (2019), s. 69–86.
- Szymański K., *Osobliwość Raymonda Kurzweila jako wizja dziejów*, „Hybris” nr 46 (2019), s. 51–68.
- Trzciniński Ł., *Koncepcje rozwoju i mitologie przyszłości*, Wydawnictwa AGH, Kraków 2016.

Panintelligence

Summary

We want to consider the problem of the gradual transformation of a universe completely devoid of intelligence into a universe that can be said to be an intelligent entity or a person. The entity can either include only its decision-making system (mind, data processing system) or the entirety, comprising both the decision-making system and its executive system, which would refer to the mechanism that enables its functioning. We are interested in the question of whether the environment external to the intelligent entity can be gradually altered and incorporated into its subjectivity, specifically within the scope of its executive system of the decision-making system. Furthermore, can the ultimate scope of its executive system coincide with the scope of the universe? This would require the entity to have knowledge about itself, which concerns, on the one hand, its own interior (in our terminology, the executive system) and, on the other hand, hypotheses extending beyond this interior that enable a holistic view of itself. Knowledge about its own interior cannot be complete, and any real increase in this knowledge can only be assessed from an “external” perspective, in the light of metaphysical self-interpretation based on hypothetical assumptions. This “metaphysics” is also a necessary premise for motivation to engage in any fully autonomous and autotelic activity.

Keywords: panintelligence, decision-making system, autotelic values, person