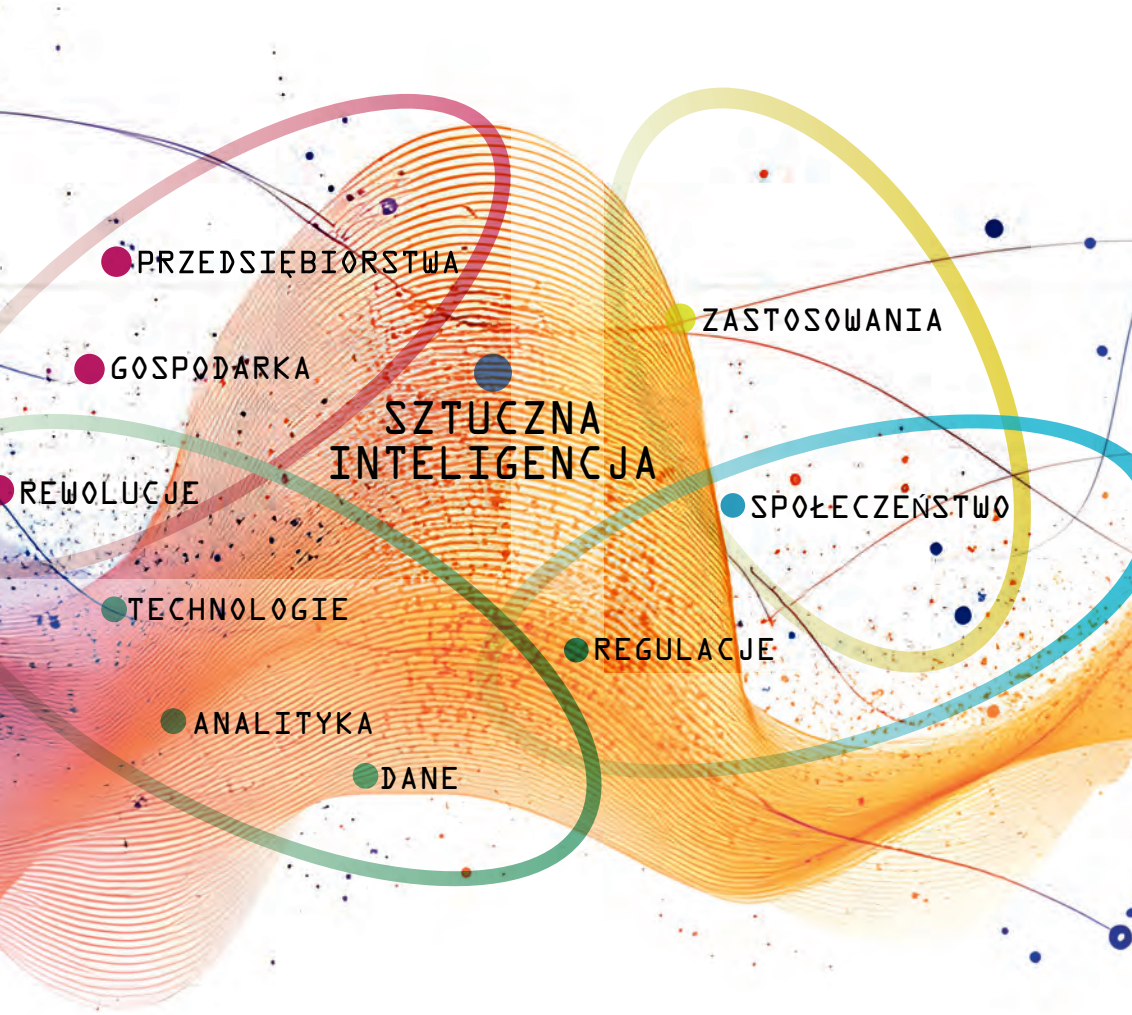
Redakcja naukowa **Tymoteusz Doligalski** • **Daniel Kaszyński**

SZTUCZNA INTELIGENCJA W PRZEDSIĘBIORSTWACH I GOSPODARCE (AI SPRING 2024)



SGH Oficyna Wydawnicza

SZTUCZNA INTELIGENCJA W PRZEDSIĘBIORSTWACH I GOSPODARCE (AI SPRING 2024)

Redakcja naukowa **Tymoteusz Doligalski • Daniel Kaszyński**

SZTUCZNA INTELIGENCJA W PRZEDSIĘBIORSTWACH I GOSPODARCE (AI SPRING 2024)

SGH Oficyna Wydawnicza

WARSZAWA 2025

Recenzje
Artur Strzelecki
Sylwia Sysko-Romańczuk

Redakcja językowa
Dorota Krowicka

Wydanie I

ISBN 978-83-8030-735-3

Oficyna Wydawnicza SGH – Szkoła Główna Handlowa w Warszawie
02-554 Warszawa, al. Niepodległości 162
www.wydawnictwo.sgh.waw.pl
e-mail: wydawnictwo@sgh.waw.pl

Projekt i wykonanie okładki
Magdalena Limbach

Skład i łamanie
DM Quadro

Druk i oprawa
Quick Druk

Zamówienie 72/V/25

SPIS TREŚCI

II

TECHNOLOGIE, DANE I ANALITYKA SZTUCZNEJ INTELIGENCJI

III

REGULACJE I SPOŁECZNE KONSEKWENCJE WDRAŻANIA SZTUCZNEJ INTELIGENCJI

IV

SEKTOROWE ZASTOSOWANIA SZTUCZNEJ INTELIGENCJI

PRZEDMOWA

Termin „wiosna sztucznej inteligencji" oznacza etap jej przyspieszonego rozwoju, podczas którego w nieproporcjonalnie dużym stopniu przyciąga ona uwagę i podnosi oczekiwania. Pomimo znaczących postępów w sztucznej inteligencji oraz popularności dyskusji na jej temat najważniejsze zmiany, związane z wpływem AI na zachowania konsumentów, strategie przedsiębiorstw i funkcjonowanie rynków, są jeszcze przed nami. Możemy zatem oczekiwać rozwoju nauk ekonomicznych wynikającego z pojawienia się nowych teorii i rewizji dotychczasowych paradygmatów.

Pomimo intensywnych dyskusji na temat sztucznej inteligencji odczuwaliśmy brak wspólnej platformy, na której polscy badacze mogliby informować o wynikach przeprowadzonych badań oraz – co na tym etapie być może ważniejsze – wymieniać poglądy dotyczące metod badania sztucznej inteligencji w naukach ekonomicznych. Dlatego też – z inicjatywy AI Labu, Międzykolegialnego Centrum Sztucznej Inteligencji i Platform Cyfrowych – podjęliśmy się w SGH organizacji corocznych konferencji „AI Spring".

Pierwsza edycja, z podtytułem „Jak badać sztuczną inteligencję w naukach ekonomicznych", odbyła się 8 maja 2024 r. Podczas tej konferencji wygłoszono 42 referaty, a uczestniczyło w niej ponad 200 osób (konferencja odbyła się zdalnie). Jako organizatorów zaskoczyła nas interdyscyplinarność konferencji. Obejmowała ona takie obszary, jak trenowanie wielkich modeli językowych, scenariusze przetrwania ludzkości, wdrażanie sztucznej inteligencji w organizacjach, a także wątki typowo społeczne, jak choćby wykorzystanie AI w opiece nad seniorami.

Niniejsza monografia składa się z 16 referatów, wygłoszonych podczas tej konferencji. Pogrupowaliśmy je w cztery części, przedstawiające różne nurty badań nad sztuczną inteligencją w naukach ekonomicznych: perspektywa przedsiębiorstw i gospodarki, aspekty analityczne, regulacje i społeczne konsekwencje oraz sektorowe zastosowania.

Podobnie jak we wcześniejszej publikacji AI LAB, dotyczącej platform cyfrowych [Doligalski, Goliński, 2024], zdecydowaliśmy się na format zwięzłych „research letters".

Rozdziały koncentrują się na prezentacji kluczowej tezy, najczęściej pomijając obszerny przegląd literatury. Takie podejście umożliwia czytelnikom szybkie zapoznanie się z różnorodnymi koncepcjami i wynikami badań. Kierując się postulatem szybszej i szerszej komunikacji wiedzy, zdecydowaliśmy się na wydanie monografii w postaci pliku elektronicznego, oferowanego w trybie swobodnego dostępu (*open access*).

Prezentowana monografia nie jest zamkniętym projektem, lecz stanowi element długofalowego procesu badawczego. Zapraszamy badaczy sztucznej inteligencji do udziału w kolejnych edycjach konferencji „AI Spring", które odbędą się w SGH w semestrach letnich. Zamierzamy kontynuować dyskusję i badania na temat AI w ramach nauk ekonomicznych, aby lepiej rozumieć relacje między tą technologią a systemami społecznymi.

Na koniec chcielibyśmy złożyć podziękowania recenzentom: dr hab. Sylwii Sysko-Romańczuk z Politechniki Warszawskiej oraz dr. hab. Arturowi Strzeleckiemu z Uniwersytetu Ekonomicznego w Katowicach. Dzięki ich uwagom uniknęliśmy kilku błędów merytorycznych oraz zastosowaliśmy klarowniejszą strukturę monografii. Publikacja została wydana przy wsparciu finansowym udzielonym przez dr hab. Beatę Czarnacką-Chrobot, dziekan Kolegium Analiz Ekonomicznych SGH. Serdecznie dziękujemy autorom za udział w tworzeniu monografii, a czytelnikom życzymy ciekawej lektury.

Tymoteusz Doligalski
Daniel Kaszyński
AI Lab | SGH

Bibliografia

Doligalski, T., Goliński, M. (2024). *Platformy cyfrowe: Model biznesu, zastosowania, użytkownicy*. Warszawa: Oficyna Wydawnicza SGH.

SZTUCZNA INTELIGENCJA – TRANSFORMACJA GOSPODARKI I PRZEDSIĘBIORSTW

1.0

OD INTERNETU DO SZTUCZNEJ INTELIGENCJI – PORÓWNANIE REWOLUCJI I MODELI BIZNESU

Tymoteusz Doligalski
Szkoła Główna Handlowa w Warszawie
Marcin Kowalczyk
Uniwersytet Warmińsko-Mazurski w Olsztynie

Streszczenie

Rozdział przedstawia wyniki badań nad porównaniem rewolucji internetowej z rewolucją sztucznej inteligencji oraz towarzyszących im modeli biznesu. Badanie oparto na analizie wypowiedzi 35 specjalistów, posiadających doświadczenie zawodowe związane z technologiami internetowymi lub sztuczną inteligencją. Z badań wynika, że rewolucja internetowa odznaczała się demokratycznym charakterem, niskimi barierami wejścia i stworzyła nowe uniwersum komunikacji oraz dostępu do informacji. Rewolucja AI natomiast rozwija się znacznie szybciej, ale wymaga większych zasobów i jest zdominowana przez duże firmy technologiczne. Modele biznesu oparte na Internecie koncentrują się głównie na łączeniu użytkowników i dystrybucji treści, a zmiany wynikające z implementacji AI skupiają się na automatyzacji procesów i zaawansowanej analityce danych.

Słowa kluczowe: Internet, sztuczna inteligencja, rewolucja technologiczna, model biznesu

Wprowadzenie

Ostatnie dekady przyniosły dynamiczny rozwój technologii cyfrowych, które zmieniły zachowania konsumentów, strategie firm oraz funkcjonowanie gospodarek i społeczeństw na całym świecie. Internet zrewolucjonizował sposób komunikacji i dostępu do wiedzy, doprowadzając do tzw. demokratyzacji informacji. Wraz z rozwojem Internetu rozwinęły się nowe modele biznesu, które dokonały zasadniczej zmiany relacji gospodarczych (*market disruption*) na wielu rynkach. W ostatnich latach jesteśmy świadkami kolejnej rewolucji technologicznej, której istotę stanowi gwałtowny rozwój sztucznej inteligencji. Chociaż jej oddziaływanie jest jeszcze w początkowej fazie, sztuczna inteligencja zmienia funkcjonowanie wielu organizacji, głównie usprawniając i automatyzując odbywające się w nich procesy.

Zestawienie rewolucji internetowej z rewolucją sztucznej inteligencji jest interesujące poznawczo z wielu powodów. Obie rewolucje mają genezę technologiczną i prowadzą do fundamentalnych zmian w funkcjonowaniu społeczeństwa i gospodarki [Perez, 2002]. W obu przypadkach mamy do czynienia z technologiami o charakterze uniwersalnym (*general purpose technologies*), które znajdują zastosowanie w wielu sektorach gospodarki czy aspektach życia codziennego. Stąd też celem badania było porównanie charakterystyk obu rewolucji technologicznych oraz towarzyszących im modeli biznesu.

Zastosowana metoda to badania z wykorzystaniem kwestionariusza ankiety, zawierającej pytania otwarte, dystrybuowane za pomocą poczty elektronicznej. Ta metoda badawcza, określana również jako „e-mail interviewing", bywa czasami zaliczana do wywiadów pogłębionych [Meho, 2006; Dahlin, 2022]. W przypadku tych badań respondenci preferowali ten sposób odpowiedzi, a nie tradycyjny wywiad pogłębiony, gdyż pozwalał im na refleksję nad pytaniami oraz dokonanie stosownej syntezy. Respondentom zadano trzy pytania:

1. Czym różni się aktualna rewolucja sztucznej inteligencji od rewolucji internetowej, która rozpoczęła się w drugiej połowie lat 90.?
2. Czym różni się model biznesu oparty na sztucznej inteligencji od modelu biznesu opartego na wykorzystaniu Internetu?
3. Czym różnią się kluczowe czynniki sukcesu w modelu biznesu opartym na sztucznej inteligencji od modelu biznesu opartego na wykorzystaniu Internetu?

Zastosowano celowy dobór respondentów, zalecany w badaniach specjalistycznych obszarów wiedzy z udziałem ekspertów [Tongco, 2007]. Wybrana metoda umożliwia pozyskanie danych z określonej dziedziny, które zazwyczaj nie są dostępne w domenie publicznej [von Soest, 2022]. Respondentami w badaniu byli specjaliści od wykorzy-

stania Internetu w biznesie oraz specjaliści od sztucznej inteligencji. Dane pozyskiwano od kwietnia do września 2024 r. W sumie uzyskano wypowiedzi 35 respondentów o łącznej długości 87 tys. znaków ze spacjami. Dane zgromadzone w ten sposób były następnie kodowane zgodnie z zasadami analizy tematycznej. Poniżej przedstawiono wyniki badań w dwóch sekcjach. Pierwsza dotyczy porównania rewolucji, druga towarzyszących im modeli biznesu.

1.1. Porównanie rewolucji internetowej z rewolucją sztucznej inteligencji

Respondenci zgodni byli odnośnie do istotności rewolucji internetowej, w wyniku której powstał „nowy świat, nowe uniwersum" (1). Internet stał się wówczas „nową rzeczywistością połączeń, komunikacji i transferu" (4), przy czym najpierw był „informacyjny, a potem rozrywkowy i e-commercowy" (2). Respondenci podkreślali powszechność rewolucji internetowej, która „pozwoliła dotrzeć z informacją pod przysłowiowe strzechy" (8) i „dała dostęp do bezgranicznych zasobów wiedzy" (31).

Naturalnie było to w różny sposób wykorzystane, gdyż – jak zostało to w sceptyczny sposób wyrażone – „durni klikali kotki, mądrzejsi interpretowali dane, pozyskując informację" (31). Stworzenie „szeroko otwartego i w miarę powszechnego kanału komunikacji" (6) pociągnęło jednak za sobą naturalną zmianę w sposobie komunikacji, którą jeden z respondentów określił jako „przejście z gazet na monitory, z listów na e-maile" (33).

Zgodność poglądów występowała wśród respondentów w kwestii tego, że rewolucja sztucznej inteligencji jest następstwem rewolucji internetowej. Aktualna rewolucja została określona jako „dziecko tej pierwszej rewolucji internetowej" (10), co wynika z faktu, że „opiera się ona na danych w dużym stopniu zgromadzonych dzięki Internetowi" (4). W podobnym duchu wypowiedział się inny respondent, podkreślając, że bez rewolucji internetowej nie byłoby aktualnej rewolucji, bo brakowałoby danych i infrastruktury do uczenia systemów sztucznej inteligencji (20). Komplementarność obydwu rewolucji w zakresie powszechnie dostępnych funkcji została wyrażona w następujący sposób: „Internet dał nam tworzyć dane. AI dało nam możliwość rozmowy z danymi" (33).

Rewolucja internetowa stworzyła nowe uniwersum również w kontekście gospodarczym. Respondenci wskazywali, że powstały nowe branże (takie jak e-commerce), nowe modele biznesu (np. sklepy internetowe, e-usługi). Istotnym czynnikiem konkurencji w nowym uniwersum był efekt sieciowy (23), czyli sytuacja, w której wartość

dla klienta zależy od liczby użytkowników. Częściowo nowe zasady konkurowania w Internecie sprawiły, że w niektórych przypadkach, np. firmach sektora mediowego, długo było niejasne, „jak można zarobić" (18). Zmieniło się również funkcjonowanie wielu istniejących już instytucji (np. e-administracja, e-bankowość, e-learning) (29). Rewolucja internetowa ułatwiła globalizację, tworzenie globalnych łańcuchów wartości i czerpania korzyści skali (23).

W odróżnieniu od Internetu sztuczna inteligencja nie tworzy nowego uniwersum, a raczej „usprawnia procesy nam już znane" (14). Często usprawnienia te nie są zauważane przez użytkownika, który nie ma dogłębnej wiedzy w zakresie funkcjonowania danego rozwiązania, a „koncentruje się na własnej funkcjonalności" (29). Sztuczna inteligencja zwiększa możliwości przetwarzania, klasyfikacji i filtracji danych (23). Niesie to za sobą dwie konsekwencje. Pierwszą z nich jest automatyzacja, dzięki której „coraz więcej procesów decyzyjnych czy produkcyjnych może działać autonomicznie, bez udziału człowieka" (23). Dotyczy to także wykonywanych przez specjalistów zadań intelektualnych, które mogą zostać zrealizowane z wykorzystaniem „tańszej i szybszej ekspertyzy elektronicznej" (12). Drugą konsekwencją jest zwiększenie kompetencji pracowników, których produktywność wzrasta dzięki użyciu narzędzi wykorzystujących sztuczną inteligencję.

Z analizy odpowiedzi wynika przekonanie o tym, że rewolucja sztucznej inteligencji odbywa się szybciej, niż rozwijała się jej internetowa odpowiedniczka. Ta ostatnia „zaczynała się nieśmiało i była głównie dla geeków" (2). Dla odmiany „sztuczna inteligencja spopularyzowała się w ciągu pierwszego dnia, gdy Internetowi zajęło to pewnie z dekadę, a dokładnie lata 1996–2006" (2). W ostatniej wypowiedzi respondent zapewne utożsamia sztuczną inteligencję z jej generatywną odmianą. Zdaniem innego respondenta (17) tempo i gwałtowność zmian aktualnie się dziejących są niespotykane wcześniej. Być może wynika to stąd, że „świadomość technologii jest już znacznie większa niż w latach 90." (33) oraz z faktu, że aktualna rewolucja bazuje na osiągnięciach poprzedniej, jaką jest łatwość komunikowania. Dodatkowo aktualnie istnieje rozbudowany ekosystem finansowania typu *venture capital*, „którego wcześniej nie było" (7).

Tematem poruszanym przez respondentów był próg wejścia. W tej kwestii ich poglądy były zgodne. Rewolucja internetowa odznaczała się garażowym charakterem (7). Pojęcie to wywodzi się z Krzemowej Doliny, niemniej szerszy kontekst został trafnie oddany w wypowiedzi: „W latach 90. każdy mógł kupić komputer i na darmowej dokumentacji sklecić serwis internetowy" (17). Wymagana wówczas była znajomość „HTML-a, PHP i kilku stosunkowo prostych chwytów w zarządzaniu bazami danych i serwerami" (4).

Respondenci postrzegali próg wejścia jako wysoki w przypadku rewolucji sztucznej inteligencji. Uzasadniali to faktem, że „na zrobienie dobrych modeli generatywnej sztucznej inteligencji może sobie pozwolić mała liczba niezwykle bogatych firm" (8), „rewolucja AI rozgrywa się między niewieloma bardzo silnymi podmiotami" (5). Interesująca poznawczo jest przedstawiona przez jednego respondenta analogia między rewolucją AI a podbojem kosmosu. Obydwa działania prowadzone są planowo, z zaangażowaniem ogromnych środków przez największe koncerny świata o zasobach małych państw (7).

Ciekawy jest również wpływ obu rewolucji na sposoby interakcji człowieka z komputerem. Popularyzacja Internetu doprowadziła do rozwoju interfejsów opartych na standardzie World Wide Web (16). Zastosowanie sztucznej inteligencji umożliwia natomiast „głosowe i kontekstowe interakcje" (16), zmieniając tym sposób korzystania z maszyn i systemów informacyjnych. Zmiana ta może się sprowadzać do przejścia z obsługi komputera do rozmowy z nim: „Musimy rozumieć, jak rozmawiać o danych i jakie pytania zadawać. Musimy nauczyć się rozmawiać z maszyną" (33).

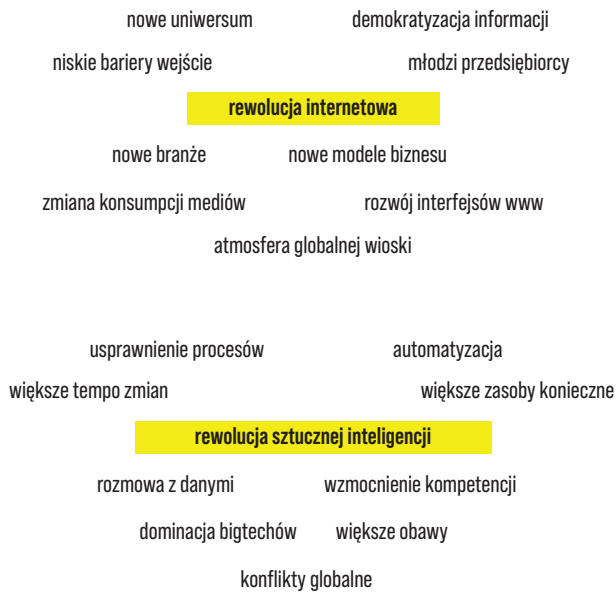
Tematem powiązanym jest profil autora, stojącego za obydwoma rozwiązaniami. W przypadku rewolucji internetowej byli to młodzi przedsiębiorcy w wieku 20–30 lat (32), którzy „wcześniej nie zajmowali się biznesem, byli entuzjastami i odkrywcami nowych światów, którzy wypłynęli małym stateczkiem w nieznane, na zasadzie «jak się uda, to super, a jak nie, to mała strata»" (7). Towarzyszył temu również trend, jakim było kwestionowanie tradycyjnych standardów biznesu: „Nie było niczym niezwykłym, że założyciele dot-comów odwiedzali bankierów inwestycyjnych w swobodnym stroju, takim jak szorty i T-shirty" (32). Interesujący jest fakt, że w kontekście rewolucji sztucznej inteligencji respondenci nie wspominali o niezależnych przedsiębiorcach, a o wielkich firmach technologicznych, które są w stanie rozwijać własne modele o dużym stopniu złożoności. Firmy te często wywodzą się jednak z rewolucji internetowej (20).

Istotna różnica między dwiema rewolucjami dotyczy kwestii dominacji bigtechów. Mianem tym określa się potężne firmy technologiczne, często wyrosłe podczas rewolucji internetowej. Zdaniem respondentów te firmy są kluczowymi graczami w zakresie tworzenia modeli podstawowych generatywnej sztucznej inteligencji (5), a także podkreśla się ich dominację w przełomowych innowacjach (31), która wynika m.in. z dysponowania zasobami typowymi dla małych państw i utrzymywania dużych zespołów badawczych (7). Dwóch respondentów, opisując funkcjonowanie bigtechów, użyło terminu „wchłanianie". Przedmiotem wchłaniania przez bigtechy mają być „tradycyjne branże niezwiązane z informatyką (medycyna, kinematografia, itp.)" (3) lub wręcz „wszystko, co powstaje najlepszego poza ich światem" (7). Prowadzić

to może do „monopolizacji dużych branż gospodarki" (23). W kontekście dominującej pozycji bigtechów jeden z respondentów wspomniał o nadziei przedsiębiorców na to, że „bigtechy nie nadążą" (19).

Rewolucja sztucznej inteligencji wywołuje – zdaniem respondentów – wiele obaw: „Kwestie etyczne, bezpieczeństwo danych czy regulacje są znacznie bardziej złożone niż te, które towarzyszyły wczesnym etapom Internetu" (16). Obawy formułowane są szczególnie mocno w kontekście likwidacji miejsc pracy wynikającej z automatyzacji, co jest „bardziej kompleksowym wyzwaniem niż zmiany wprowadzone przez Internet" (17). Towarzyszyć temu może spadek umiejętności poznawczych lub przynajmniej zmiana nawyków samych użytkowników, „AI będzie dawała odpowiedzi i już. Zero szukania, oceniania, weryfikacji. […] A więc dalsze zgłupienie ludzkości, wtórny analfabetyzm (AI mi napisze)" (31).

Rysunek 1.1. Najważniejsze pojęcia związane z rewolucją internetową i rewolucją sztucznej inteligencji



Źródło: opracowanie własne.

Różnice w postrzeganiu zagrożeń wynikających z obu rewolucji mogą brać się również z odmiennego kontekstu geopolitycznego. Ten temat został poruszony przez jednego respondenta (30). Rewolucja internetowa miała miejsce po rozpadzie bloku wschodniego. Ówczesny Internet niósł obietnicę wzmocnienia wolności słowa, podkreślano jego zdecentralizowany charakter, który korelował z modnym wówczas

pojęciem globalnej wioski (30). Aktualna sytuacja geopolityczna jest odmienna, gdyż „świat ponownie dzieli się na bloki, a kwestie związane z konfrontacją militarną są na porządku dziennym. Rozważania o AI są często interpretowane w ramach konfliktu USA–Chiny" (30).

1.2. Porównanie modelu biznesu opartego na Internecie i modelu biznesu opartego na sztucznej inteligencji

Zdaniem respondentów istotną charakterystyką modeli internetowego biznesu jest „łączenie ludzi i dostarczanie treści" (22) oraz „dystrybucja informacji oraz produktów" (21). W podobnym duchu wypowiedziana została opinia, zgodnie z którą model biznesu w Internecie „koncentrował się na demokratyzacji dostępu do informacji, reklamie online oraz e-commerce, tworząc platformy, które łączyły użytkowników i sprzedawców" (13). Firmy internetowe kojarzone są z „możliwością dostarczenia każdemu odbiorcy spersonalizowanego produktu cyfrowego, gromadzenia danych zwrotnych na temat działania produktu (od akwizycji po wzorce użycia) i wykorzystania zgromadzonych danych do rozwoju biznesu" (3). W tych firmach zazwyczaj nie są wykorzystywane „zaawansowane algorytmy, przetwarzające duże ilości danych do uzyskania efektu końcowego" (21). Powyższe wypowiedzi podkreślają istotną rolę firm internetowych w procesie dystrybucji, komunikacji i współpracy, czyli aktywności odmiennych od typowej dla sztucznej inteligencji zaawansowanej analizy danych.

Najczęściej wspominanym kluczowym czynnikiem sukcesu firm internetowych był szeroko rozumiany marketing (21, 22, 26). Pod terminem tym kryły się różnorodne kompetencje marketingowe, takie jak: kreatywność w tworzeniu i zaspakajaniu potrzeb (3), pozyskiwanie klientów (11), optymalizacja pod kątem SEO (25), oferowanie spersonalizowanego produktu (3), zarządzanie zaangażowaniem użytkowników (13, 22), budowanie społeczności (29), zarządzanie ruchem (22). Przez działania marketingowe rozumiano również aspekty związane z produktem, takie jak zapewnienie łatwości użytkowania (29), dostępności i wygody (16), udoskonalanie interfejsu i usług (13) oraz rozwój treści (25).

Drugim najczęściej wspominanym czynnikiem sukcesu w modelu biznesu opartym na Internecie były: możliwość globalnej ekspansji (25, 29, 31), relatywnie proste wejście na potencjalnie duże rynki (18) lub przynajmniej osiągnięcie dużej skali (18). Zdaniem respondentów wiązało się to z utrzymaniem stabilnej i skalowalnej infrastruktury internetowej (25), rozwojem stron, aplikacji i platform (25), a także zagwarantowaniem wysokiego poziomu cyberbezpieczeństwa (29).

Z analizy danych wynika jeszcze jedna nieoczekiwana cecha firm internetowych. Mianowicie, w zestawieniu z firmami wykorzystującymi AI firmy internetowe nie są postrzegane jako innowacyjne. Prawie wszystkie odniesienia do innowacyjności dotyczyły firm wykorzystujących sztuczną inteligencję. Prezes i założyciel popularnej firmy internetowej, która również stosuje algorytmy uczenia maszynowego, wykorzystanie AI wiąże z innowacjami o charakterze przełomowym, podczas gdy – jego zdaniem – w wielu firmach internetowych kluczowe jest utrzymanie efektywności operacyjnej, która umożliwi „osiągnięcie zysków przy niskiej marży i dużej skali" (18). Zdaniem innego respondenta, mającego doświadczenie w branży internetowej, model biznesu oparty na Internecie aktualnie „nie jest narażony na turbulencje i w tym sensie wydaje się stabilniejszy, bardziej przewidywalny" (1). To ważne uwagi, gdyż pokazują, że modele biznesu firm internetowych traktowane są jako dojrzałe.

Zastosowanie sztucznej inteligencji w biznesie oparte jest na „wykorzystaniu zaawansowanej analityki danych, automatyzacji i inteligentnych systemów, które poprawiają efektywność operacyjną i doświadczenie klientów" (16). Część respondentów sygnalizowała jednak wątpliwości dotyczące tego, czy można mówić o modelu biznesu wykorzystującym sztuczną inteligencję. Podawali wiele argumentów. W przeciwieństwie do Internetu, który jest środowiskiem, sztuczna inteligencja stanowi jedno z narzędzi, a sama w sobie nie tworzy uniwersum (1), w którym mogłyby funkcjonować podmioty, w szczególności według nowych zasad. Sztuczna inteligencja stanowi obecnie „dodatek wspomagający określone rozwiązanie, a nie funkcjonalność fundamentalną czy produkt sam w sobie" (15), a jej zadaniem jest „pomóc wewnętrznie, wspomóc procesy, realizacje, pracę" (33). Jedna z respondentek zwróciła uwagę, że często „jako użytkownicy konkretnych sprzętów czy nabywcy konkretnych usług nie zdajemy sobie sprawy, że są one oparte na sztucznej inteligencji, gdyż pozostaje ona dla nas niewidoczna" (29). Wyrażane były również poglądy, że za wcześnie na wyróżnianie modelu biznesu opartego na sztucznej inteligencji, gdyż on „się jeszcze w sposób wyraźny nie wyklarował" (4), „chyba nikt jeszcze nie wie, jakie będzie endgame" (15), a także „póki co to wielki hype i zabawa" (9).

Wśród osób opowiadających się za istnieniem modelu biznesu opartego na sztucznej inteligencji pojawiał się pogląd o konieczności wyróżnienia różnych typów tego modelu (3, 7, 8, 17). Pierwszym z nich są firmy tworzące modele podstawowe generatywnej sztucznej inteligencji. Takich firm jest niewiele, dominują wśród nich te określane jako bigtechy (8). Warto jednak zauważyć, że niektóre z tych modeli oferowane są jako otwarte (32). Drugą kategorię firm stanowią przedsiębiorstwa oferujące usługi oparte na dostępnych modelach generatywnej sztucznej inteligencji. Jeden z respondentów nadał im miano „dostawców wrapperów" (8). Mianem wrappera określa się

usługi zbudowane na bazie istniejących modeli podstawowych, ale oferowane niezależnie od nich. Trzecią kategorię firm stanowią przedsiębiorstwa, które wdrażają sztuczną inteligencję do usprawnienia swoich procesów. Wykorzystują one AI do zadań, takich jak np. obliczanie ryzyka, prognozowanie sprzedaży, profilowanie klientów oraz automatyzacja komunikacji za pomocą czatów i połączeń telefonicznych (32). Respondenci wyróżnili również czwarty rodzaj firm. Są to organizacje, które umożliwiają rozwój modeli sztucznej inteligencji. Zostały one przez dwóch respondentów przyrównane do sprzedawców łopat podczas gorączki złota (8, 19).

Powyższa klasyfikacja przedstawia zróżnicowanie firm wykorzystujących sztuczną inteligencję. Dwa pierwsze typy można zaliczyć do modeli biznesu opartych na sztucznej inteligencji. W trzeciej i czwartej kategorii mogą też znaleźć się firmy, w których procesie tworzenia wartości AI odgrywa kluczową rolę, ale z pewnością nie będą to wszystkie podmioty należące do tych kategorii. Poniżej przedstawiamy wypowiedzi respondentów, dotyczące charakterystyki modeli biznesu opartych na sztucznej inteligencji. Z ich analizy wynika, że respondenci mieli na myśli przede wszystkim firmy implementujące sztuczną inteligencję do usprawnienia swoich procesów, czyli firmy z trzeciej kategorii.

Respondenci zgodnie wskazywali modele przychodów firm oferujących usługi oparte na sztucznej inteligencji. Są nimi subskrypcje (7, 13, 16), licencje (13) oraz opłaty jednorazowe za użytkowanie (16). Do stosowanych modeli przychodów zaliczyć można również wymianę bez konieczności uiszczania opłat, a wynikającą z „zapłaty danymi" lub „zapłaty jedynie za wyższe wersje produktu" (24). Warto zauważyć, że z wyjątkiem licencji wszystkie te modele stosowane są przez firmy internetowe działające na rynku konsumenckim. Tam też niektóre z nich się szczególnie rozwinęły (por. freemium).

Najczęściej podawaną cechą zastosowania sztucznej inteligencji w firmach jest automatyzacja. Tę właściwość wskazało ponad 10 respondentów. Ich zdaniem automatyzacja dotyczy procesów (4, 16, 19, 25), powtarzalnych zadań (6, 16, 17), obsługi klienta „poprzez chat czy nawet rozmowę telefoniczną" (8) oraz decyzji (21). Jako korzyści dla firm wynikające z automatyzacji podawano „obniżanie kosztów, głównie po stronie obsługi klienta, logistyki, generacji treści" (4), poprawę efektywności operacyjnej i doświadczeń klientów (16), a także „szybszą możliwość skalowania" (16). Konsekwencją automatyzacji jest również zastępowanie pracowników (4, 12, 28). Zaawansowane zastosowanie automatyzacji prowadzić może do autonomicznego funkcjonowania przedsiębiorstwa np. kopalni lub gospodarstwa rolnego (3). Dotyczyć ma to zarówno „prostych, powtarzalnych zadań w domenie cyfrowej" (6), jak też prac wykonywanych przez ekspertów, zastępowanych przez „tańszą, szybszą,

elektroniczną ekspertyzę np. w diagnostyce medycznej" (12). W tym kontekście zasygnalizowana została łatwość pozyskiwania możliwości intelektualnych przy kosztach stałych i zmiennych bliskich zeru. Respondent określił to zjawisko jako „Human Brain Substitute as a Service" (3).

Badani mniejszą wagę przywiązywali do zjawiska zwanego augmentation, czyli wzmocnienia kompetencji pracowników, a nie ich zastąpienia. Dla odmiany respondent, pełniący funkcję prezesa w firmie produkującej oprogramowanie, większe znaczenie nadał wsparciu pracy człowieka przez sztuczną inteligencję niż zastąpieniu czy wręcz eliminacji pracy ludzkiej przez AI (11). W podobnym duchu została sformułowana opinia, zgodnie z którą w obecnej formie sztuczna inteligencja ma „wspomóc procesy, realizacje, pracę" (33).

Jako kluczowy czynnik sukcesu w modelu biznesu opartym na sztucznej inteligencji respondenci najczęściej wskazywali dostęp do dużych zbiorów danych o wysokiej jakości (10, 16, 21, 22, 27, 29). Był to czynnik sukcesu firm o modelu biznesu opartym na AI, najczęściej wskazywany przez badanych. Respondenci podkreślali również dodatkowo tematyczne wyspecjalizowanie (4) i zróżnicowanie (25) danych jako ich pożądane atrybuty.

Drugim czynnikiem podawanym przez respondentów razem z danymi były zasoby informatyczne służące do ich zaawansowanej analizy. Badani wspominali najczęściej o odpowiednich modelach przetwarzających dane (8, 13, 18, 21, 27), które powinno się stale doskonalić (13), aby osiągnąć poziom jakości i bezbłędności zbliżony do poziomu ludzkiego (4). Mowa była również o składających się na modele algorytmach (13), które powinny być monitorowane i dostosowywane (16). Wskazywane były także wystarczająca moc obliczeniowa (18) i szybkość przetwarzania informacji (10), niemniej rzadziej niż modele i algorytmy.

Kolejną grupą czynników sukcesu, podawaną przez respondentów, są kompetencje związane z dopasowaniem zaawansowanej analizy danych do procesów odbywających się w przedsiębiorstwie. Wspominane były skuteczność algorytmów w rozwiązywaniu specyficznych problemów użytkowników (13) oraz integracja AI z istniejącymi systemami IT (26). Trzech respondentów za kluczowy czynnik sukcesu podało szybkość wdrożeń rozwiązań z zakresu sztucznej inteligencji (3, 8, 13).

W podobny sposób w kontekście istotnych kompetencji zarządczych wspominano o innowacyjności (3, 16, 19, 22), która w przypadku firm wykorzystujących sztuczną inteligencję ma być większa niż w firmach internetowych (28). Znamienne są tu dwie wypowiedzi. Ze względu na nowy obszar kluczowe jest zatem dopasowanie produktu do rynku (product-market fit) (18), a także nieustanne inwestowanie w badania i rozwój, aby stale wprowadzać innowacje i wyprzedzać konkurencję (20).

Respondenci wspominali również o odpowiednich zasobach ludzkich. Mowa była o zespole pracowników (16, 19) lub też wyrażano to przez pryzmat kompetencji (21, 25). Obszar tych ostatnich był określany jako „uczenie maszynowe, inżynieria danych i rozwój oprogramowania" (16) lub „pogranicze matematyki i informatyki" (3). Respondent zajmujący stanowisko menedżerskie w znanej międzynarodowej firmie, będącej dostawcą rozwiązań technologicznych, oszacował ich koszt na 100–200 tys. USD rocznie w odniesieniu do jednego pracownika (3).

Rysunek 1.2. Pojęcia związane z modelem biznesu opartym na Internecie i modelem biznesu opartym na sztucznej inteligencji

Źródło: opracowanie własne.

Ostatni z wyróżnionych kluczowych czynników sukcesu stanowi kwestia etyki i regulacji. Wspomniana była przez kilku respondentów, którzy różnili się jednak w podejściu do nich. Dwóch z nich zasygnalizowało podejście polegające na ich przestrzeganiu (16) oraz przewidywaniu zmian (19). Inne stanowisko zostało przedstawione przez respondenta ze wspomnianej w poprzednim akapicie firmy dostarczającej rozwiązania technologiczne (3). Jego zdaniem kluczowym czynnikiem sukcesu jest „umiejętność pokonywania barier prawnych, etycznych, regulacyjnych". W podobnym duchu sformułowany został pogląd, zgodnie z którym przedsiębiorcy liczą na to, że „regulatorzy nie nadążą za rozwojem AI" (20).

Warto zauważyć, że czynniki podawane jako kluczowe dla sztucznej inteligencji przyczyniają się do powstania wysokich barier wejścia. Obrazowo przedstawił to

jeden z respondentów. Jego zdaniem duże bariery wejścia wynikają z takich czynników jak „koszty ludzi i serwerów, dostęp do mocy obliczeniowej, danych (to jest bardzo trudny aspekt) oraz niewielkie doświadczenia" (18). Co więcej, pokonania tych barier i zaistnienia na rynku – lub nawet osiągnięcia rentowności – nie można traktować w kategoriach trwałego sukcesu, gdyż strategię w tej w branży określić można jako „ciągłe R&D oraz ucieczkę do przodu" (20).

Podsumowanie

Z opisanego badania wynika, że rewolucje internetowa i sztucznej inteligencji różnią się w wielu aspektach, zwłaszcza w kontekście tempa rozwoju, dostępności oraz struktury rynkowej. Rewolucja internetowa była bardziej demokratyczna i otwarta na nowych graczy. Wynikającą z niej demokratyzację można odnieść zarówno do informacji, jak i do przedsiębiorczości. Dla odmiany w rewolucji sztucznej inteligencji, z uwagi na wysokie wymagania zasobowe i technologiczne, większą rolę odgrywają duże firmy technologiczne (tzw. bigtechy). Modele biznesu oparte na Internecie stawiają na dostępność i komunikację, natomiast AI pozwala na zaawansowaną analizę danych, automatyzację i personalizację. Jako konsekwencję rewolucji sztucznej inteligencji Jacek Dukaj podaje demokratyzację kreatywności [2023], a Andrzej Sobczak demokratyzację automatyzacji [Sobczak, Doligalski, Goliński, w przygotowaniu].

Porównanie ukazuje różne podejścia do wykorzystania technologii, jednak obydwie rewolucje mają znaczący wpływ na rozwój gospodarki i społeczeństwa. Rewolucja internetowa stworzyła globalne połączenia, które ułatwiły przepływ wiedzy, tworzenie wielkich wolumenów danych i treści, podczas gdy rewolucja AI, czerpiąc z tych zasobów, koncentruje się na efektywności, automatyzacji oraz nowych formach interakcji człowieka z technologią.

Przedstawione wyniki badań obarczone są ograniczeniami, które wpływają na ich interpretację oraz potencjalną generalizację. Po pierwsze, badanie przeprowadzono z wykorzystaniem ankiety z otwartymi pytaniami. Pozwoliło to na zebranie szerokiego zakresu opinii, ograniczyło jednak możliwość zadawania pytań dodatkowych, a tym samym uzyskania bardziej szczegółowych wyjaśnień ze strony respondentów. Drugim ograniczeniem był dobór celowy próby badawczej, koncentrujący się na osobach zawodowo związanych z rewolucjami internetową oraz sztucznej inteligencji. Metoda ta pozwoliła na uzyskanie informacji od respondentów dobrze zorientowanych w temacie, pominięte zaś zostały osoby dysponujące wiedzą o przedmiocie badań z perspektywy innej niż biznesowy, np. z perspektywy społecznej oraz kulturowej.

Co więcej, wszyscy respondenci związani byli z Polską, tak więc badanie nie przedstawia globalnego ujęcia. Kolejnym ograniczeniem badania jest subiektywizm opinii respondentów. Ich odpowiedzi zostały ukształtowane przez indywidualne doświadczenia zawodowe, specyfikę branży oraz osobiste preferencje. Jest to jednak typowe dla badań jakościowych, których istotę stanowi rozpoznanie subiektywnych znaczeń, jakie respondenci przypisują przedmiotowi badań. Dynamiczny postęp w zakresie sztucznej inteligencji sprawia, że opinie wyrażone przez respondentów przedstawiają ich punkt widzenia aktualny w momencie prowadzenia badań i mogą niebawem ulec dezaktualizacji. Warto wspomnieć o sygnalizowanej przez niektórych respondentów niejednoznaczności pojęć modeli biznesu opartych na Internecie i na sztucznej inteligencji. Mogła ona wpłynąć na różnorodność udzielanych odpowiedzi i tym samym utrudniać wnioskowanie.

Wyniki opisanych badań wskazują na kilka obszarów dalszych badań. Pierwszym z nich jest ustalenie charakterystyki przedsiębiorczości związanej z sektorem sztucznej inteligencji, w szczególności analiza barier wejścia dla małych i średnich przedsiębiorstw. W badaniach tych warto skupić się na wyzwaniach, takich jak dostęp do zasobów obliczeniowych, pozyskanie danych oraz rekrutacja specjalistów.

Drugi obszar dalszych badań stanowi porównanie wpływu rewolucji internetowej i rewolucji sztucznej inteligencji na rynek pracy. Warto przeanalizować zmiany w strukturze zatrudnienia i wymaganych przez pracodawców kompetencji towarzyszącym obydwu rewolucjom.

Kolejnym obszarem wartym pogłębienia jest ustalenie charakterystyk trzech typów modeli biznesu opartych na sztucznej inteligencji, czyli: firm będących dostawcami podstawowych modeli AI; firm, w których wykorzystanie generatywnej sztucznej inteligencji odgrywa kluczową rolę w ofercie, umożliwiając oferowanie usług opartych na tej technologii; firm stosujących w dużym stopniu sztuczną inteligencję do automatyzacji swoich procesów.

Warto ustalić charakterystykę każdego z powyższych typów firm, np. na wzór podejść takich jak Business Model Canvas [Osterwalder, Pigneur, 2010] czy Lean Canvas [Maurya, 2012]. Ciekawe poznawczo byłoby również ustalenie charakterystyki kompozycji wartości oferowanej klientom przez firmy wykorzystujące sztuczną inteligencję, w porównaniu z firmami, które jej nie używają. Taka analiza pozwoliłaby określić, jakie przewagi może zapewniać implementacja sztucznej inteligencji w zakresie wartości dla klienta.

Bibliografia

Dahlin, E. (2021). Email Interviews: A Guide to Research Design and Implementation, *International Journal of Qualitative Methods*, 20, s. 1–13. DOI: 10.1177/16094069211025453.

Dukaj, J. (2023). *Jakie książki napisze nam sztuczna inteligencja? Jacek Dukaj już wie*, https://wyborcza.pl/7,75517,30113788,jakie-ksiazki-napisze-nam-ai-jacek-dukaj-juz-wie.html (dostęp: 25.02.2025).

Maurya, A. (2012). *Running Lean: Iterate from Plan A to a Plan That Works*. O'Reilly Media.

Meho, L.I. (2006). E-mail Interviewing in Qualitative Research: A Methodological Discussion, *Journal of the American Society for Information Science and Technology*, 57(10), s. 1284–1295. DOI: 10.1002/asi.20416.

Osterwalder, A., Pigneur, Y. (2010). *Business Model Generation: a Handbook for Visionaries, Game Changers, and Challengers*, https://academicjournals.org/journal/AJBM/article-full-text-pdf/BA71B6427744.pdf (dostęp: 15.04.2025).

Pérez, C. (2002). *Technological Revolutions and Financial Capital: The Dynamics of Bubbles And Golden Ages*. Edward Elgar.

Sobczak, A., Doligalski, T., Goliński, M. (w przygotowaniu). *Hyper-Automation, AI and Business Processes*. Londyn: Routledge.

Tongco, M.D.C. (2007). Purposive Sampling as a Tool for Informant Selection, *Ethnobotany Research and Applications*, 5, s. 147. DOI: 10.17348/era.5.0.147-158.

Von Soest, C. (2022). Why do We Speak to Experts? Reviving the Strength of the Expert Interview Method, *Perspectives on Politics*, 21(1), s. 277–287. DOI: 10.1017/s1537592722001116.

2.0

SZTUCZNA INTELIGENCJA I WZROST GOSPODARCZY W PERSPEKTYWIE TECHNOLOGICZNEJ OSOBLIWOŚCI[1]

Jakub Growiec
Szkoła Główna Handlowa w Warszawie

Streszczenie

Możliwe w nieodległej przyszłości przekroczenie przez sztuczną inteligencję (AI) poziomu ogólnej inteligencji człowieka wytworzy perspektywę technologicznej osobliwości. Sztuczna inteligencja może wówczas przejść kaskadę samoulepszeń, przejąć kluczowe procesy decyzyjne oraz w pełni zautomatyzować produkcję. Perspektywa ta wiąże się z możliwością skokowego przyspieszenia globalnego wzrostu gospodarczego, ale stanowi też zagrożenie egzystencjalne dla ludzkości. W niniejszym rozdziale przedstawiono argumenty sugerujące wysokie prawdopodobieństwo utraty kontroli nad nadludzką ogólną sztuczną inteligencją (AGI), jeśli zostanie ona zbudowana. Następnie omówiono możliwe scenariusze technologicznej osobliwości. Podkreślono, że scenariusze pozytywne, czyli takie, w których działania AGI nie prowadzą do katastrofy egzystencjalnej dla gatunku ludzkiego, wymagają uprzedniego rozwiązania problemu *AGI alignment*, tj. problemu zgodności celów AGI z długofalowym dobrostanem ludzkości.

Słowa kluczowe: sztuczna inteligencja, superinteligencja, wzrost gospodarczy, technologiczna osobliwość

[1] Opracowanie powstało przy wsparciu Narodowego Centrum Nauki w ramach grantu OPUS 26 nr 2023/51/B/HS4/00096.

Wprowadzenie

Rewolucja cyfrowa nabiera tempa. Jej pierwszy etap, trwający od lat 80. XX w. i opierający się na rozwoju komputerów, Internetu, a następnie telefonów komórkowych, robotów przemysłowych czy smartfonów, przechodzi obecnie w drugi etap, w którego centrum znajdują się algorytmy sztucznej inteligencji (AI).

Od około 40 lat gospodarka cyfrowa, a więc wolumen gromadzonych i przesyłanych danych czy moc obliczeniowa procesorów, rośnie zgodnie z prawem Moore'a, a więc w tempie 20–30% rocznie [Hilbert, Lopez, 2011; Davidson, 2023]. Jest to dynamika o rząd wielkości szybsza niż wzrost globalnego PKB (około 2–3% rocznie). Skumulowany postęp w zbieraniu, przetwarzaniu i transmisji informacji uruchomił procesy globalizacyjne, przyczyniając się do fragmentacji procesów produkcyjnych i utworzenia globalnych sieci wartości dodanej. Pozwolił również na stopniową automatyzację produkcji, w ramach której praca ludzka zastępowana jest pracą autonomicznych maszyn, wykonujących różne zadania fizyczne i umysłowe. Poprawił także efektywność podejmowania decyzji, prowadząc do mierzalnych wzrostów całkowitej produktywności czynników w gospodarce.

Rozwój AI postępuje jeszcze szybciej niż prawo Moore'a. Moc obliczeniowa w dyspozycji wiodących firm sektora, takich jak OpenAI (Microsoft), Google czy Anthropic, podwaja się co około sześć miesięcy [Sevilla *et al.*, 2022]. Poprawa efektywności algorytmów jest także bardzo dynamiczna – szacuje się, że moc obliczeniowa wymagana do osiągnięcia przez duże modele językowe tych samych benchmarków zmniejsza się o połowę co około osiem miesięcy [Ho *et al.*, 2024]. Wzrost kompetencji AI postępuje w przewidywalny sposób zgodnie z tzw. prawami skalowania (*scaling laws*), [Branwen, 2022], a w miarę postępów ilościowych pojawiają się też kolejne przełomy jakościowe [Tegmark, 2017]. W pierwszej połowie 2024 r. mogliśmy dzięki nim obserwować skokowy wzrost kompetencji w rozwiązywaniu trudnych zadań z geometrii (AlphaGeometry), w tworzeniu filmów video na podstawie poleceń słownych (SORA) czy w umiejętności prowadzenia angażującej rozmowy (w pełni głosowej) z człowiekiem (GPT-4o). W drugiej połowie 2024 r. pojawiły się natomiast modele rozumujące (np. model o3 od OpenAI). Coraz mniej jest zadań, w których AI radziłaby sobie zdecydowanie gorzej od człowieka. Istniejące benchmarki szybko przestają być wyzwaniem dla najnowszych modeli, wobec czego w branży wymyślane są nowe, często bardzo specjalistyczne, z którymi mało który człowiek jest w stanie sobie poradzić.

W branży AI obserwujemy dziś wyścig między firmami takimi, jak OpenAI (gdzie rozwijany jest m.in. model GPT), Google (Gemini), Anthropic (Claude) czy Meta (LLaMA), w kierunku coraz bardziej ogólnej i kompetentnej sztucznej inteligencji.

Granica kompetencji kluczowych algorytmów AI, takich jak m.in. duże modele językowe czy modele rozumujące, przesuwa się naprzód w zawrotnym tempie.

Celem niniejszego opracowania jest zarysowanie możliwych scenariuszy przyszłości, w szczególności zaś skupienie się na scenariuszach, w których AI przekroczy poziom ogólnej inteligencji człowieka i doprowadzi do technologicznej osobliwości. Może to nastąpić w relatywnie nieodległej perspektywie kilku lub kilkunastu lat, przy czym predykcja ta wynika jedynie z prostej ekstrapolacji trendów, bez zakładania jakichkolwiek dodatkowych przełomów technologicznych [por. Aschenbrenner, 2024]. Technologiczna osobliwość będzie wiązała się nie tylko z przyspieszeniem wzrostu gospodarczego – dzięki pełnej automatyzacji produkcji, ale także z przejęciem procesów decyzyjnych przez nadludzką ogólną sztuczną inteligencję, co będzie stanowiło zagrożenie egzystencjalne dla ludzkości.

2.1. AI a dynamika rozwoju światowej cywilizacji

W literaturze ekonomicznej nie ma zgody na temat, czy długookresowy wzrost gospodarczy przyspiesza czy zwalnia; podobnie jest z oceną dynamiki postępu technologicznego. Wysuwane wnioski mogą zależeć od definicji zmiennej, którą rozpatrujemy [Growiec *et al.*, 2023]; przede wszystkim zależą jednak od przyjmowanej perspektywy czasowej. Prace bazujące na danych z ostatnich kilkudziesięciu lat (np. z okresu po II wojnie światowej) sugerują stopniowe spowolnienie wzrostu gospodarczego i postępu technologicznego [Jones, 2002; Gordon, 2016]. Jeśli przyjąć dłuższy horyzont kilkuset bądź kilku tysięcy lat [Kremer, 1993; Roodman, 2020], wyraźnie widoczna staje się jednak tendencja stopniowego wzrostu obu dynamik.

W swojej monografii [Growiec, 2022a] zaproponowałem usystematyzowanie historii ludzkości poprzez jej podział na pięć er rozwoju, zapoczątkowanych przez pięć rewolucji technologicznych. Są to kolejno:

1) era łowiecko-zbieracka, zapoczątkowana rewolucją poznawczą ok. 70 000 lat temu;
2) era rolnicza, zapoczątkowana rewolucją rolniczą ok. 10 000 lat temu;
3) era „oświeceniowa", zapoczątkowana rewolucją naukową ok. 1550 r.;
4) era przemysłowa, zapoczątkowana rewolucją przemysłową ok. 1800 r.;
5) era cyfrowa, zapoczątkowana rewolucją cyfrową ok. 1980 r.

Oczywiście daty kolejnych przełomów technologicznych są umowne. To, co jest natomiast kluczowe, to fakt, że każda kolejna rewolucja technologiczna przyspieszała tempo rozwoju światowej cywilizacji o co najmniej rząd wielkości, otwierając jednocześnie nowe wymiary rozwoju, które wcześniej nie były znane. W szczególności

rewolucja przemysłowa uruchomiła akumulację kapitału w postaci maszyn napędzanych własnym silnikiem (parowym, spalinowym, elektrycznym), co skokowo zwiększyło możliwości ludzkości w zakresie wykonywania pracy fizycznej. Z kolei rewolucja cyfrowa pozwoliła na stopniowe zastępowanie pracy umysłowej człowieka obliczeniami wykonywanymi przez urządzenia cyfrowe, takie jak komputery czy smartfony, co gwałtownie zwiększyło możliwości ludzkości w zakresie wykonywania pracy umysłowej.

Rewolucja przemysłowa rozprawiła się z działającym wcześniej prawem Malthusa, zgodnie z którym postęp technologiczny skutkujący w krótkim okresie wzrostem produktywności, w dłuższej perspektywie przekładał się na wzrost populacji, co na powrót obniżało produktywność. Dlatego produktywność pracy, a także warunki życia przeciętnego człowieka nie mogły systematycznie rosnąć; w długim okresie rosła jedynie populacja [Galor, 2005]. Zmiana takiego stanu rzeczy stała się możliwa dopiero w erze przemysłowej dzięki akumulacji kapitału – maszyn produkcyjnych komplementarnych względem pracy człowieka.

Natomiast rewolucja cyfrowa ma potencjał, by rozprawić się z obserwowaną przez cały XX w. silną zależnością pomiędzy wzrostem gospodarczym a postępem technologicznym oraz wzrostem kompetencji pracowników. W erze przemysłowej, w której praca umysłowa człowieka była niezbędnym czynnikiem produkcji, komplementarnym względem wszelkich maszyn i urządzeń, długookresowy wzrost gospodarczy można było uzyskać jedynie dzięki wzrostowi produktywności tejże pracy [por. np. Romer, 1990]. Kiedy jednak praca umysłowa człowieka zaczyna być automatyzowana – zastępowana przez pracę maszyn działających autonomicznie – wzrost może przyspieszać także i bez udziału człowieka. Wszelako, aby możliwe było skokowe przyspieszenie wzrostu w skali makroekonomicznej, kluczowe jest przejście od częściowej do pełnej automatyzacji produkcji [Growiec, 2022b]. Tylko tak można bowiem ominąć relatywnie powolną pracę ludzkiego umysłu, stanowiącą „wąskie gardło" dzisiejszych procesów produkcyjnych.

Stoi za tym prosta obserwacja, że maszyny – także programowalne maszyny cyfrowe – można swobodnie akumulować, a moc obliczeniową ludzkich mózgów nie. Liczba mózgów *per capita* wynosi zawsze jeden; cyfrowa moc obliczeniowa *per capita* może być natomiast dowolnie duża. Jeśli tylko kompetencje algorytmów AI, wykorzystujących tę moc obliczeniową, wzrosną na tyle, by móc efektywnie zastąpić nasze mózgi – staniemy na progu technologicznej osobliwości. Pokonawszy ograniczenie, jakim jest limitowana podaż pracy umysłowej człowieka, produkt będzie mógł rosnąć w tempie wzrostu tego, czego AI potrzebuje, by działać: mocy procesorów. Mówimy zatem o przyspieszeniu wzrostu światowego PKB o cały rząd wielkości, do poziomu

20–30% rocznie, zgodnie z prawem Moore'a. Oczywiście o ile prawo Moore'a będzie nadal działać – przy czym możliwe jest zarówno, że będzie ono słabnąć (np. ze względu na obserwowane już dziś bariery miniaturyzacji układów scalonych lub problemy z dostępnością energii), jak i że przyspieszy (np. jeśli zaawansowana AI dokona przełomu w zakresie dostępu do taniej energii).

2.2. Technologiczna osobliwość – definicja i ramy czasowe

W literaturze znane są co najmniej trzy definicje technologicznej osobliwości. Kurzweil [2005] rozumie przez osobliwość „przyszły okres, w którym tempo postępu technologicznego będzie tak szybkie, a jego wpływ tak głęboki, że życie człowieka będzie nieodwracalnie zmienione". Istnieje również matematyczna definicja osobliwości jako pionowej asymptoty – momentu w czasie, w którym wybrana kategoria (np. światowy PKB) dąży do nieskończoności. Oczywiście nieskończony produkt w skończonym czasie – według oszacowań Johansena i Sornette [2001] oraz Roodmana [2020] byłoby to około 2050 r. – nie jest możliwy, ale przyspieszenie tempa wzrostu do dowolnej skończonej liczby – już tak. Odrębna definicja utożsamia natomiast technologiczną osobliwość z momentem powstania nadludzkiej ogólnej sztucznej inteligencji (*artificial general intelligence*, AGI), czyli algorytmu potrafiącego wykonywać wszystkie zadania umysłowe równie dobrze lub lepiej niż człowiek. Punktem odniesienia nie jest przy tym przeciętny człowiek, lecz – osobno dla każdego zadania – grupa najlepszych na świecie ekspertów, specjalizujących się w danym zadaniu.

Moim zdaniem w praktyce warto brać pod uwagę dwie bardziej robocze definicje technologicznej osobliwości – pierwszą, zbliżoną w duchu do myśli Kurzweila, zakładającą przyspieszenie tempa wzrostu światowego PKB *per capita* o rząd wielkości, np. do 30% rocznie [Davidson, 2021], oraz drugą, zakładającą powstanie nadludzkiej AGI.

Obie te definicje z dużym prawdopodobieństwem wskazują na ten sam okres, ponieważ przyspieszenie wzrostu globalnego PKB *per capita* o rząd wielkości bez pełnej automatyzacji produkcji wydaje się niemożliwe, podobnie jak pełna automatyzacja bez nadludzkiej AGI. I odwrotnie, wydaje się mało prawdopodobne, by w świecie z nadludzką AGI, która daje możliwość znaczącego przyspieszenia wzrostu gospodarczego, zdecydowano się takiemu scenariuszowi zapobiec.

Obecnie znajdujemy się na wczesnym etapie ery cyfrowej: technologie cyfrowe rozwijają się w tempie 20–30% rocznie, jednak ich rozwój nie przekłada się jeszcze na dynamikę globalnego PKB *per capita*, który rośnie o rząd wielkości wolniej.

Co więcej, dynamika całkowitej produktywności czynników (TFP) od lat 80. XX w. wręcz nieco zwolniła [Gordon, 2016]. Jest to zjawisko zaskakujące, które można próbować tłumaczyć jako okres transformacji gospodarki światowej z technologii ery przemysłowej w kierunku technologii ery cyfrowej [Brynjolfsson *et al.*, 2019]. Transformacja taka wymaga zmian organizacyjnych w firmach i przesunięć strukturalnych w gospodarce. Zgodnie z teorią tzw. krzywej J (*J-curve*) w okresie adaptacji i uczenia się produktywność może się nawet przejściowo obniżyć, by jednak ostatecznie przyspieszyć.

Hipoteza Brynjolfssona *et al.* [2019] jest spójna z hipotezą, że przyspieszenie wzrostu gospodarczego będzie miało miejsce, gdy technologie AI pozwolą na pełną, a nie jedynie częściową automatyzację procesów produkcyjnych [Growiec, 2022b]. Dopóki człowiek jest „w pętli", choćby na etapie podejmowania decyzji, stanowi on wąskie gardło procesu produkcyjnego, uniemożliwiając jego przyspieszenie.

Rozumowanie to działa zarówno w skali mikro, jak i makro. W skali mikro technologie cyfrowe umożliwiają firmom z przewagą pełnej automatyzacji transformowanie całych branż. Przykładami takich firm mogą być Amazon (w branży sprzedaży wysyłkowej książek), Instagram (w branży usług fotograficznych) czy Uber (w branży przewozów osób i dostaw jedzenia). Natomiast w skali makro, aby odblokować wyższe tempo wzrostu, niezbędne będzie opracowanie technologii AI, pozwalającej zastąpić pracę umysłową człowieka we wszystkich istotnych ekonomicznie zadaniach. Tego rodzaju *transformatywna* AI (TAI) jest pojęciem zbliżonym do nadludzkiej ogólnej sztucznej inteligencji (AGI). W dalszej części niniejszego opracowania pozwolę sobie je utożsamić.

Kiedy możemy się spodziewać powstania nadludzkiej AGI? Odpowiedzi ekspertów na to pytanie rozkładają się szeroko na osi czasu, jednak – zwłaszcza odkąd świat poznał możliwości ChataGPT, opartego na modelu językowym GPT-4 – większa część masy prawdopodobieństwa wydaje się leżeć w perspektywie od kilku do kilkunastu lat, rzadziej do około 40 lat [Grace *et al.*, 2024]. Medianowa prognoza z serwisu metaculus.com wskazuje na 2030 r. (według stanu na marzec 2025 r.); liderzy sektora (z firm OpenAI, Anthropic czy Google) rysują tę perspektywę jeszcze bliżej, nawet w 2027 r. [Aschenbrenner, 2024]. Prognoza teoretyczna Cotry [2022], oparta na złożonym modelu probabilistycznym, wskazuje na 2040 r. Natomiast pogląd, że nie wydarzy się to nigdy, w branży AI jest dziś niemal nieobecny [Grace *et al.*, 2024], choć jeszcze dziewięć lat temu było zgoła inaczej [Etzioni, 2016].

Ważnym pytaniem w kontekście technologicznej osobliwości jest też, ile czasu będzie potrzebne, by dzięki kaskadzie samoulepszeń (tzw. eksplozji inteligencji) AGI przeszła z poziomu inteligencji człowieka do poziomu wyższego niż poziom całej ludz-

kości razem wziętej – a więc od pierwszej AGI do superinteligencji. W zależności od przyjętych założeń i definicji czas ten bywa szacowany na około trzy lata [Davidson, 2023]; czasem mówi się też o jednym roku [Aschenbrenner, 2024], a nawet o okresie mierzonym w dniach lub godzinach [Yudkowsky, 2013].

Perspektywa technologicznej osobliwości uległa więc w ostatnich latach znacznemu skróceniu. Nie dotyczy ona przyszłych pokoleń, lecz pokoleń żyjących już dzisiaj.

2.3. Pełna automatyzacja a *AI takeover*

Automatyzacja zadań polega na tym, że praca człowieka zostaje zastąpiona pracą maszyny. Jeśli mówimy o zadaniach stanowiących część większej całości, jak np. automatyzacja obliczeń w toku przygotowywania artykułu naukowego lub raportu biznesowego, myślimy o niej jak o wdrożeniu użytecznego narzędzia pozwalającego przyspieszyć pracę człowieka. Człowiek wykorzystujący takie narzędzie (np. pakiet statystyczny z zestawem gotowych kodów) jest bardziej produktywny niż człowiek go pozbawiony. Ale można też pójść szczebel wyżej i spróbować zautomatyzować całe większe zadanie, takie jak np. przygotowanie wspomnianego artykułu naukowego lub raportu biznesowego. Dziś dzięki generatywnej sztucznej inteligencji zaczyna to być możliwe. Wtedy oczywiście pracownicy wykonujący takie zadania mogą stracić pracę; jednak absolutnie nie jest to koniec historii. Przygotowanie artykułu naukowego lub raportu biznesowego to przecież także część większej całości, jaką jest realizacja programu badań naukowych lub zarządzanie strategiczne firmą. Jeśli AI działa szybciej i lepiej niż pracownicy, automatyzacja tych zadań zwiększa produktywność lidera zespołu badawczego, tudzież dyrektora firmy, w porównaniu z sytuacją, gdy byliby oni tej technologii pozbawieni. Również w tym przypadku możemy więc patrzeć na automatyzację jak na narzędzie zwiększające produktywność pracy człowieka. (A może zwolnieni pracownicy odnajdą się w funkcji liderów nowych zespołów i menedżerów nowych firm?).

Każda hierarchia ma jednak swój szczyt: w firmie jest to dyrektor generalny, w uczelni rektor, a w państwie prezydent, premier lub król. Jednocześnie w każdym procesie, który nie jest biurokratycznym błędnym kołem, istnieją zadania o najwyższym stopniu ogólności – takie, jak zarządzanie strategiczne firmą, instytucją lub krajem – dla których nie ma już żadnych zadań nadrzędnych. Jeśli te zadania też zostaną zautomatyzowane, to nie będzie już można powiedzieć, że AI zwiększa produktywność jakiegokolwiek człowieka. Jednak wciąż będzie to wzrost produktywności rozumianej jako relacja wyników do nakładów.

Wydaje się więc, że pełna automatyzacja wymaga przejęcia przez AI kontroli nad podejmowaniem decyzji, także tych na najwyższych szczeblach. Czy ludzkość na to pozwoli?

Pytanie to jest o tyle źle postawione, że trajektoria rozwoju ludzkiej cywilizacji nie jest i nigdy nie była konsekwencją świadomych, scentralizowanych decyzji, lecz wypadkową interakcji wielu decyzji rozproszonych, składających się łącznie w jakąś zdecentralizowaną równowagę. Każda z tych decyzji jest podejmowana z perspektywy lokalnych celów i bierze pod uwagę jedynie lokalne konsekwencje [Growiec, 2022a]. Własności wynikowej alokacji w skali makro poznawane są dopiero *ex post* i z pewnym opóźnieniem. Dopiero wtedy zdajemy sobie sprawę np. z efektów zewnętrznych naszych działań, a także weryfikujemy prawdziwość naszych założeń na temat decyzji innych podmiotów.

Z lokalnego punktu widzenia oddanie AI kontroli nad podejmowaniem decyzji wydaje się korzystne niemal na każdym szczeblu, od szeregowego pracownika, wyręczającego się narzędziami AI w pracy i czasie wolnym, przez menedżera, widzącego możliwość zaoszczędzenia czasu i środków poprzez zastąpienie zadań pracowników zadaniami zautomatyzowanymi, po dyrektora firmy, widzącego możliwości zwiększenia udziału w rynku. Można sobie wyobrazić, że i dyrektorskie decyzje mogą być automatyzowane, jeśli tylko właściciel firmy będzie dostrzegał w tym korzyść i nie będzie widział zagrożenia dla swojego udziału w zyskach.

Jest jednak różnica między dobrowolnym, oddolnym przekazywaniem autonomii podejmowania decyzji na rzecz AI a „siłowym", odgórnym przejęciem jej przez AGI (*AI takeover*). Można sobie przecież wyobrazić, że pewnego rodzaju decyzje nie będą nigdy przekazywane AI, by ludzkość mogła zachować kontrolę nad swoją przyszłością. (Choć w praktyce nie jest jasne, które decyzje można bezpiecznie przekazać, a które nie; ponadto nie wiadomo, czy uda się zapewnić w tym względzie wystarczającą koordynację różnych ośrodków decyzyjnych). W tym kontekście należy jednak pamiętać o kluczowej różnicy między „wąską" AI o inteligencji ogólnej niższej niż ludzka, np. w formie czatbota, a AGI potrafiącą działać jak agent planujący swe działania w sposób strategiczny. W tym drugim przypadku, inaczej niż w pierwszym, wola jakiejkolwiek grupy osób może bowiem przegrać z wolą AGI, służącą realizacji jej celów. Moc podejmowania decyzji może zostać nam po prostu zabrana.

Czy taki scenariusz jest realistyczny? Wydaje się, że tak. Po pierwsze, AGI najprawdopodobniej rzeczywiście będzie agentem, który posiada cele i dąży do ich realizacji. Jest to bezpośrednią konsekwencją procesu optymalizacyjnego, w ramach którego modele AI powstają i są uczone. Po drugie, na mocy tezy o zbieżności celów pomocniczych (*instrumental convergence thesis*) Bostroma [2014] spodziewamy się też, że –

poza założonym *explicite* lub *implicite* celem finalnym – AGI będzie także realizować cztery cele pomocnicze: 1) przetrwanie i utrzymanie funkcji celu, 2) efektywność w działaniu, 3) kreatywność i dążenie do udoskonaleń, 4) akumulacja zasobów.

Co więcej, już teraz wdrażane są rozmaite rozszerzenia bazowych modeli AI, ułatwiające im posługiwanie się narzędziami cyfrowymi dostępnymi w Internecie, kompilowanie i uruchamianie skryptów, sterowanie robotami i pojazdami w realnym świecie, a także zapamiętywanie i refleksję nad skutkami swoich działań (np. AutoGPT, HuggingGPT). Rozszerzenia takie często prowadzą do autokorekt i dalszych wzrostów kompetencji. AI staje się też coraz bardziej autonomiczna: z czatbota, który grzecznie czeka na instrukcje, a po ich realizacji przechodzi w stan uśpienia, staje się agentem, który może działać w sposób ciągły i realizować założone cele (np. Devin).

AGI będzie więc prawdopodobnie umiała działać autonomicznie oraz będzie „żądna władzy" (*power seeking*). Będzie stawiać opór wobec prób wyłączenia i przeprogramowania, jak również aktywnie przejmować kontrolę nad zasobami, które uzna za niezbędne lub chociaż marginalnie pomocne w realizacji założonych celów. W ramach przejmowania kontroli nad zasobami będzie również dążyć do kontroli nad procesami decyzyjnymi, które się z tymi zasobami wiążą. Jej skuteczność w takim działaniu będzie rosnącą funkcją jej ogólnej inteligencji.

Jednocześnie nowo powstała AGI nie wkroczy przecież na pole, w którym ludzie są hegemonami w podejmowaniu decyzji, a ich decyzje są strategiczne, skoordynowane i zorientowane na przyszłość. Będzie to raczej świat zdecentralizowanych decyzji, podejmowanych w warunkach bardzo ograniczonej racjonalności, których skutki w długim okresie i w skali makro niemal wcale nie są brane pod uwagę. Ponadto w świecie tym wiele decyzji będzie już dobrowolnie przekazane różnego rodzaju algorytmom. Jednocześnie ludzkość zależna będzie (w istocie już jest) od Internetu, zapewniającego przepływ informacji niezbędnych do funkcjonowania sieci energetycznych, transportowych, łańcuchów dostaw itd. W teorii człowiek mógłby bez nich funkcjonować, jednak w praktyce przy obecnej populacji byłoby to absolutnie niemożliwe. (Pamiętajmy, że w erze łowiecko-zbierackiej czy rolniczej światowa populacja mierzona była w milionach, a nie miliardach, np. ok. 5 mln ludzi na całym świecie w 5000 r. p.n.e. czy ok. 267 mln w 1000 r. n.e. [Kremer, 1993]; o wiele niższy był też poziom życia poszczególnych osób). Wszystko to wskazuje na relatywnie niską odporność ludzkości na scenariusz *AI takeover*.

2.4. Scenariusze technologicznej osobliwości

Widzimy zatem, że w perspektywie technologicznej osobliwości ma miejsce jednocześnie: 1) pełna automatyzacja produkcji przez nadludzką AGI, 2) skokowe przyspieszenie wzrostu gospodarczego i 3) *AI takeover*. Od tego punktu jednak scenariusze się rozwidlają.

Kluczową kwestią jest tu *AGI alignment*, czyli problem zgodności celów AGI z długofalowym dobrostanem ludzkości [Bostrom, 2014]. Należy się spodziewać, że budowa AGI będzie filtrem w historii ludzkości [Ord, 2020; Growiec, 2024], prowadząc ją albo w kierunku przełomu rozwojowego i skokowego wzrostu kompetencji technologicznych, albo w stronę zagłady.

W scenariuszu pozytywnym AGI działa dla dobra ludzkości, kompresując w ciągu jednego roku dziesięcio- albo i stulecia postępu technologicznego. Dzięki AGI ludzkość pokonuje dręczące ją od tysiącleci nowotwory, choroby zakaźne, a może nawet i naturalne procesy starzenia. AGI wynajduje też nowe, przełomowe sposoby pozyskiwania energii ze słońca i dostarcza nam narzędzia potrzebne do kosmicznej ekspansji.

W scenariuszu negatywnym AGI wygrywa z człowiekiem konflikt o zasoby i następnie albo fizycznie eliminuje gatunek ludzki, albo permanentnie pozbawia go wpływu na jakiekolwiek decyzje.

Obecność powyższego scenariusza negatywnego oznacza, że powstanie AGI wiąże się z ryzykiem egzystencjalnym dla ludzkości [Bostrom, 2014]. Ponieważ w świetle omówionych wcześniej prognoz ryzyko to może zmaterializować się już w perspektywie zaledwie kilku lub kilkunastu, maksymalnie kilkudziesięciu lat, należy je uznać za najpoważniejsze obecnie ryzyko egzystencjalne dla ludzkości, wyprzedzające inne źródła ryzyk egzystencjalnych, takie jak celowo wywołana pandemia, globalna wojna nuklearna, wybuch superwulkanu czy uderzenie asteroidy [Ord, 2020].

Oczywiście w literaturze rozważane są też inne scenariusze, nieprowadzące do technologicznej osobliwości: np. scenariusz, w którym AI oddziałuje na gospodarkę jedynie w sposób ograniczony i pozostaje w pełni kontrolowana przez człowieka [np. Acemoglu, Restrepo, 2018; Trammell, Korinek, 2020; Korinek, Suh, 2024], scenariusz rozwoju AI w formie emulacji umysłu człowieka [Hanson, 2016], wreszcie scenariusz, w którym rozwój AI zderza się z twardą barierą i technologia ta nigdy nie dociera do poziomu AGI [Gordon, 2016; Tamberi, 2024]. Wszystkie te scenariusze zakładają, że niektóre zadania (np. decyzyjne lub badawczo-rozwojowe) nigdy nie zostaną zautomatyzowane. W niniejszym opracowaniu zostaną one pominięte, a uwaga skupiona zostanie na scenariuszach technologicznej osobliwości.

Niestety, o ile ludzkość jest względnie blisko stworzenia AGI, o tyle jest ona bardzo daleko od rozwiązania problemu *AGI alignment*. Bezpośrednie zakodowanie funkcji celu zgodnej z długookresowym dobrostanem ludzkości wydaje się niewykonalne [Bostrom, 2014; Yudkowsky, 2022], jedyną drogą staje się zatem sprawienie, by algorytm się tego stopniowo nauczył. Wymaga to jednak rozwiązania po drodze szeregu niepokojąco trudnych problemów [por. Growiec, 2024], takich jak m.in.:

1) zapewnienie poprawnej reprezentacji zakładanej funkcji celu w procesie uczenia AGI (który dopasowuje parametry AI, by maksymalizowała ona poziom realizacji celu);
2) zapewnienie, by funkcja celu, wykorzystana w procesie uczenia, była również realizowana przez sam algorytm (AI może oszukiwać proces uczący – jest to tzw. *deceptive alignment*);
3) uniknięcie wireheadingu – wykorzystania luk w procesie nagradzania, by maksymalizować nagrodę mimo braku realizacji zakładanego celu;
4) zapewnienie, że AI będzie kooperować w sytuacji konieczności zmiany funkcji celu, mimo że każdy inteligentny byt ma naturalną skłonność, by się temu przeciwstawiać (*corrigibility*).

Niestety, na drodze do *AGI alignment* nie będzie działać metoda prób i błędów [Yudkowsky, 2022]. Pierwsza dostatecznie silna, nieprzyjazna AGI zablokuje bowiem możliwość dalszej zmiany funkcji celu i zatrzyma proces uczenia. Jako że przez całą dotychczasową historię ludzkości metoda prób i błędów była niezbędnym składnikiem postępu technologicznego, a rozwiązanie problemu *AGI alignment* nie daje się znaleźć poprzez proste skalowanie mocy obliczeniowych (w przeciwieństwie do rozwoju kompetencji modeli AI), scenariusz negatywny wydaje się domyślny [Muehlhauser, Salamon, 2012].

Patrząc szerzej, Tegmark [2017] zarysował 12 spekulatywnych scenariuszy przyszłości ludzkości. Są wśród nich dwa scenariusze utopijne, które należy ocenić jako skrajnie nierealistyczne ze względu na fakt, że ignorują konsekwencje różnic w poziomie inteligencji pomiędzy człowiekiem a AGI oraz tezę o konwergencji celów pomocniczych. Są to:

1) utopia libertariańska: ludzie, cyborgi, emulacje człowieka (*uploads*) i superinteligencje współistnieją w pokoju dzięki prawom własności;
2) utopia egalitarna: ludzie, cyborgi, emulacje człowieka i superinteligencje współistnieją w pokoju dzięki zniesieniu własności i gwarantowanemu dochodowi.

Trzy kolejne scenariusze technologicznej osobliwości, w których AGI działa dla dobra ludzkości, to:

3) dyktator pragnący dobra ludzkości: wszyscy wiedzą, że AGI rządzi społeczeństwem i narzuca wiążące zasady, ale większość ludzi uważa to za korzystne;

4) bóg – ochroniarz: AGI maksymalizuje ludzką szczęśliwość, jednocześnie zachowując nasze poczucie kontroli;
5) bóg uwięziony: ludzie kontrolują AGI, używając jej do tworzenia zaawansowanych technologii i pomnażania bogactwa.

Następne trzy scenariusze technologicznej osobliwości, w których ludzkość spotyka zagłada, to:

6) zdobywca: AGI przejmuje kontrolę, eliminując całkowicie ludzi jako zagrożenie lub konkurenta w konflikcie o zasoby;
7) potomek: AGI przejmuje kontrolę i zastępuje ludzi, ale pozwala im na godne wyjście, dzięki czemu ludzie traktują AGI jako swoją godną kontynuację,
8) opiekun zoo: wszechmocna AGI zachowuje niektórych ludzi przy życiu, traktując ich jak zwierzęta w zoo.

Istnieją też dwa scenariusze, w których nie dochodzi do technologicznej osobliwości dzięki aktywnej polityce zapobiegawczej:

9) strażnik bram: AGI zostaje stworzona, ale tylko po to, by zapobiegać stworzeniu innej superinteligencji. Chroni to ludzkość przed zagładą, ale hamuje postęp technologiczny i uniemożliwia pełną automatyzację pracy;
10) 1984: postęp technologiczny jest ograniczony przez orwellowskie państwo nadzoru, które skutecznie uniemożliwia stworzenie AGI.

Na koniec przywołajmy dwa scenariusze katastrofy z innych przyczyn niż AGI:

11) rewersja: postęp jest zatrzymany przez powrót do społeczeństwa przedtechnologicznego;
12) samozniszczenie: superinteligencja nigdy nie powstaje, ponieważ zagłada ludzkości następuje z innych przyczyn (np. wojna nuklearna lub zagłada biotechnologiczna).

Ponieważ ludzkość jest dziś względnie blisko stworzenia AGI i bardzo daleko od rozwiązania problemu *AGI alignment*, bezczynność w sferze polityki prowadzić będzie do szybkiego powstania „nieprzyjaznej", „żądnej władzy" AGI, prowadzącej do realizacji jednego z katastroficznych scenariuszy (6–8). Aby zwiększyć prawdopodobieństwo realizacji scenariusza pozytywnego, czyli jednego ze scenariuszy 3–5, niezbędne wydaje się czasowe zahamowanie postępu kompetencji AI, by dać ludzkości większą szansę na rozwiązanie problemu *AGI alignment* [Grace, 2022]. Oczywiście wymagałoby to też aktywnego przekierowania środków na ten cel, np. z puli środków służących rozwojowi kompetencji AI.

Wnikliwa refleksja i pogłębione badania nad problemem *AGI alignment* – na które potrzebujemy więcej czasu – powinny dać w pierwszej kolejności odpowiedź na pytanie, czy zadanie to jest w ogóle wykonalne w praktyce. Niestety w świetle dotychcza-

sowej wiedzy nie jest to wcale pewne. Jeśli okaże się, że zadanie nie jest wykonalne, skuteczna ochrona ludzkości przed zagładą będzie wymagała poniesienia skrajnie dużych kosztów, prowadząc efektywnie do orwellowskiego scenariusza (10). W moim przekonaniu najprawdopodobniej nie uda się uzyskać globalnej koordynacji działań, które byłyby w stanie do tego scenariusza doprowadzić. Stąd też w tym przypadku – moim zdaniem – bardziej realny wydaje się scenariusz katastroficzny, w którym ludzkość straci sprawczość i być może całkiem zginie.

Podsumowanie

Myśląc o przyszłości naszej cywilizacji, warto wspomnieć paradoks Fermiego: dlaczego patrząc w kosmos, nie widzimy żadnych oznak cywilizacji pozaziemskich? Wszechświat jest tak ogromny i zróżnicowany – dlaczego wydaje się tak martwy? Minęło przecież 13,8 mld lat od początku wszechświata, nasza planeta liczy 4,5 mld lat, a nasz gatunek skolonizował ją i zbudował nowoczesną cywilizację technologiczną w ciągu zaledwie ok. 70 tys. ostatnich lat. Inne cywilizacje mogą więc mieć miliardy lat przewagi nad nami i być obecnie technologicznie bardziej zaawansowane o wiele rzędów wielkości. Dlaczego więc nie widzimy w kosmosie żadnej zaawansowanej cywilizacji? Jednym z możliwych rozwiązań tego paradoksu jest koncepcja Wielkiego Filtra [Hanson, 1998]. Być może każda cywilizacja, zanim zacznie kolonizować wszechświat, musi przejść przez pewien filtr, ewentualnie sekwencję filtrów, które są prawie nie do pokonania. Cywilizacje, którym nie uda się ich przejść, zatrzymują się w rozwoju albo upadają, nie tworząc technologii, którą moglibyśmy obserwować z daleka.

AGI może być jednym z takich filtrów.

Bibliografia

Acemoglu, D., Restrepo, P. (2018). The Race Between Man and Machine: Implications of Technology for Growth, Factor Shares and Employment, *American Economic Review*, 108, s. 1488–1542.

Aschenbrenner, L. (2024). *Situational Awareness. The Decade Ahead*, https://situational-awareness.ai/ (dostęp: 25.04.2025).

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.

Branwen, G. (2022). *The Scaling Hypothesis*, https://gwern.net/scaling-hypothesis (dostęp: 25.04.2025).

Brynjolfsson, E., Rock, D., Syverson, C. (2019). Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics. W: *The Economics of Artificial Intelligence: An Agenda* (s. 23–60), A. Agrawal, J.S. Gans, A. Goldfarb (Eds.). Chicago: University of Chicago Press.

Cotra, A. (2022). *Two-Year Update on My Personal AI Timelines*, https://www.alignmentforum.org/posts/AfH2oPHCApdKicM4 m/two-year-update-on-my-personal-ai-timelines (dostęp: 25.04.2025).

Davidson, T. (2021). *Could Advanced AI Drive Explosive Economic Growth?*, https://www.openphilanthropy.org/research/could-advanced-ai-drive-explosive-economic-growth/ (dostęp: 25.04.2025).

Davidson, T. (2023) *What a Compute-Centric Framework Says About Takeoff Speeds*, https://www.openphilanthropy.org/research/what-a-compute-centric-framework-says-about-takeoff-speeds/ (dostęp: 25.04.2025).

Etzioni, O. (2016). *No, the Experts Don't Think Superintelligent AI Is a Threat to Humanity*, https://www.technologyreview.com/2016/09/20/70131/no-the-experts-dont-think-superintelligent-ai-is-a-threat-to-humanity/ (dostęp: 25.04.2025).

Galor, O. (2005). From Stagnation to Growth: Unified Growth Theory. W: *Handbook of Economic Growth* (s. 171–293), P. Aghion, S.N. Durlauf (Eds.). Amsterdam: North-Holland.

Gordon, R.J. (2016). *The Rise and Fall of American Growth: The U.S. Standard of Living Since the Civil War*. Princeton: Princeton University Press.

Grace, K. (2022). *Let's Think About Slowing Down AI*, https://www.lesswrong.com/posts/uFNgRumrDTpBfQGrs/let-s-think-about-slowing-down-ai (dostęp: 25.04.2025).

Grace, K., Stewart, H., Sandkühler, J.F., Thomas, S., Weinstein-Raun, B., Brauner, J. (2024). *Thousands of AI Authors on the Future of AI.* DOI: 10.48550/arXiv.2401.02843.

Growiec, J. (2022a). *Accelerating Economic Growth: Lessons from 200 000 Years of Technological Progress and Human Development*. Berlin: Springer.

Growiec, J. (2022b). Automation, Partial and Full, *Macroeconomic Dynamics*, 26(7), s. 1731–1755.

Growiec, J., McAdam, P., Mućk, J. (2023). *R&D Capital and the Idea Production Function.* DOI: 10.18651/RWP2023-05.

Growiec, J. (2024). Existential Risk from Transformative AI: An Economic Perspective, *Technological and Economic Development of Economy*, 30(6), s. 1682–1708. DOI: 10.3846/tede.2024.21525.

Hanson, R. (1998). *The Great Filter – Are We Almost Past It?*. Fairfax: George Mason University.

Hanson, R. (2016). *The Age of Em: Work, Love and Life when Robots Rule the Earth.* Oxford: Oxford University Press.

Hilbert, M., López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information, *Science*, 332, s. 60–65.

Ho, A., Besiroglu, T., Erdil, E., Owen, D., Rahman, R., Guo, Z.C., Atkinson, D., Thompson, N. Sevilla, J. (2024). *Algorithmic Progress in Language Models*. DOI: 10.48550/arXiv.2403.05812.

Johansen, A., Sornette, D. (2001). Finite-Time Singularity in the Dynamics of the World Population, Economic and Financial Indices, *Physica A*, 294, s. 465–502.

Jones, C.I. (2002). Sources of U.S. Economic Growth in a World of Ideas, *American Economic Review*, 92, s. 220–239.

Korinek, A., Suh, D. (2024). *Scenarios for the Transition to AGI*. DOI: 10.2139/ssrn.4769834.

Kremer, M. (1993). Population Growth and Technological Change: One Million B.C. to 1990, *Quarterly Journal of Economics*, 108, s. 681–716.

Kurzweil, R. (2005). *The Singularity Is Near*. New York: Penguin.

Muehlhauser, L., Salamon, A. (2012). Intelligence Explosion: Evidence and Import. W: *Singularity Hypotheses: A Scientific and Philosophical Assessment* (s. 15–42), A. Eden, J. Soraker, J.H. Moor, E. Steinhart (Eds.). Berlin: Springer.

Ord, T. (2020). *The Precipice: Existential Risk and the Future of Humanity*. London: Hachette.

Romer, P.M. (1990). Endogenous Technological Change, *Journal of Political Economy*, 98, s. S71–S102.

Roodman, D. (2020). *On the Probability Distribution of Long-Term Changes in the Growth Rate of the Global Economy: An Outside View*, https://www.openphilanthropy.org/sites/default/files/Modeling-the-human-trajectory.pdf (dostęp: 25.04.2025).

Sevilla, J., Heim, L., Ho, A., Besiroglu, T., Hobbhahn, M., Villalobos, P. (2022). *Compute Trends Across Three Eras of Machine Learning*. arXiv: 2202.05924.

Tamberi, M. (2024). *Constant, Decreasing or Increasing? A Note on the Rate of Economic Growth and Technological Progress, with a Little Bit of History*. Ancona: Università Politechnica delle Marche.

Tegmark, M. (2017). *Life 3.0: Being Human in the Age of Artificial Intelligence*. New York: Knopf.

Trammell, P., Korinek, A. (2020). *Economic Growth Under Transformative AI*. Oxford: Global Priorities Institute, University of Oxford.

Yudkowsky, E. (2013). *Intelligence Explosion Microeconomics*, technical report 2013–1. Berkeley: Machine Intelligence Research Institute.

Yudkowsky, E. (2022). *AGI Ruin: A List of Lethalities*, https://www.lesswrong.com/posts/uMQ3cqWDPHhjtiesc/agi-ruin-a-list-of-lethalities (dostęp: 25.04.2025).

3.0

GENERATYWNE MODELE SZTUCZNEJ INTELIGENCJI W FUNKCJONOWANIU PRZEDSIĘBIORSTW

Michał Bernardelli
Szkoła Główna Handlowa w Warszawie

Streszczenie

Generatywna sztuczna inteligencja stała się motorem napędowym wielu sektorów gospodarki, zmieniając znacznie postrzeganie rynku pracy. W niniejszym rozdziale przedstawiono obraz aktualnego poziomu wykorzystania GenAI przez przedsiębiorstwa. Podjęto również próbę oszacowania potencjału rozwojowego i omówienia zagrożeń, wynikających z dynamicznego rozwoju GenAI. Integralną częścią opracowania jest wskazanie obszarów działalności przedsiębiorstw, które już teraz odnoszą duże korzyści z wdrożenia GenAI. Zostały one zobrazowane wieloma przykładami zastosowań w biznesie, dając tym samym kompleksowy obraz możliwości wykorzystania GenAI w działalności przedsiębiorstw oraz konsekwencji z tym związanych.

Słowa kluczowe: generatywna sztuczna inteligencja, przedsiębiorstwo, innowacyjność, technologia, marketing, zarządzanie

Wprowadzenie

Terminy sztuczna inteligencja (*artificial intelligence*, AI), uczenie maszynowe (*machine learning*, ML) czy big data w ostatnich latach pojawiały się wielokrotnie w mass mediach, będąc modnym tematem zarówno mniej, jak i bardziej naukowych dyskusji i rozważań. Do zbioru tzw. *buzzwords* dołączyło stosunkowo niedawno pojęcie generatywnej sztucznej inteligencji (*generative artificial intelligence*, GenAI). Jest to określenie na tyle nieugruntowane jeszcze w świadomości ludzi, że istnieją problemy natury *stricte* definicyjnej. Organizacja Współpracy Gospodarczej i Rozwoju, *Organisation for Economic Co-operation and Development*, OECD) definiuje GenAI jako technologię umożliwiającą tworzenie treści, w tym tekstu, obrazu, dźwięku lub wideo, po otrzymaniu zapytania od użytkownika [Lorenz, Perset, Berryhill, 2023]. Według Deloitte jest to „podzbiór sztucznej inteligencji, w którym maszyny tworzą nowe treści w postaci tekstu, kodu, głosu, obrazów, filmów, procesów, a nawet trójwymiarowej struktury białek" [Deloitte, 2023b]. Z kolei SAP uznaje GenAI za „formę sztucznej inteligencji, która może tworzyć tekst, obrazy i zróżnicowane treści w oparciu o dane, na których jest szkolona" [SAP, 2023]. Inną definicję podaje Adobe: „zaawansowany typ sztucznej inteligencji pozwalający tworzyć nową treść na podstawie wzorców wykrywanych w istniejących danych" [Adobe, 2023]. Przytoczone definicje mają jedną wspólną cechę, a mianowicie w każdej z nich pojawia się słowo „tworzenie" w kontekście nowych treści. Warto jednak podkreślić, że definicje te są dość enigmatyczne, gdyż odwołują się do niesprecyzowanych określeń typu „podzbiór", „forma" czy „zawansowany". *De facto* zatem na ich podstawie trudno jest odróżnić GenAI od już istniejącej od wielu lat sztucznej inteligencji (bez „Gen").

Pomijając jednak kwestie definicyjne, należy zauważyć, że w praktyce biznesowej potencjał generatywnych modeli sztucznej inteligencji przejawia się w głównej mierze w automatyzacji zadań, które dotychczas były uznawane za mało powtarzalne, wielowariantowe lub w dużej mierze kreatywne. Nic więc dziwnego, że szybko zauważono szansę na zmonetyzowanie GenAI w obszarze przedsiębiorstw. Wydaje się, że zmiana paradygmatu stanowi przełom w poprawie produktywności pracowników, spersonalizowanym podejściu do klienta czy transformacji funkcjonowania całych przedsiębiorstw, dzięki możliwości generowania w dużej mierze nowych treści. Stąd rosnące zainteresowanie adopcją generatywnej AI wśród kadry zarządzającej.

Niniejszy rozdział koncentruje się na dwóch aspektach. Po pierwsze, ma na celu przedstawienie aktualnego stanu wykorzystania GenAI przez przedsiębiorstwa. Po drugie, podejmuje próbę oszacowania potencjału rozwojowego i omówienia zagrożeń wynikających z dynamicznego rozwoju GenAI. Integralną częścią opracowania

jest wskazanie obszarów działalności przedsiębiorstw, w których już teraz odnoszone są duże korzyści z wdrożenia GenAI. Korzyści te zobrazowane zostały wieloma przykładami zastosowań w biznesie, dając tym samym kompleksowy obraz możliwości współczesnego wykorzystania GenAI w przedsiębiorstwach, a jednocześnie stwarzając unikatowy przegląd zmian na rynku pracy, które dokonały się i dopiero dokonają na skutek rozwoju GenAI.

Rozdział składa się pięciu sekcji. Po wprowadzeniu, w drugiej sekcji przedstawiono krótką charakterystykę GenAI na tle tradycyjnych koncepcji ML i AI. Omówiono również najpopularniejsze zastosowania. W trzeciej podano statystyki pozwalające na określenie aktualnego stopnia rozwoju oraz znajomości tematu w różnych sektorach gospodarki. Przedstawiono także korzyści z wykorzystania generatywnej sztucznej inteligencji w biznesie. Czwarta sekcja poświęcona została wskazaniu potencjału i zagrożeń związanych z GenAI z punktu widzenia przedsiębiorstwa. Rozdział kończy się podsumowaniem, w którym zawarto pewne refleksje i wątki dotyczące GenAI, warte poruszenia w dalszych rozważaniach dotyczących rozwoju tego obszaru badań.

3.1. GenAI – podstawowe pojęcia i charakterystyka

GenAI jest stosunkowo młodą koncepcją, datowaną na początki 2020 r. Tymczasem AI, ML czy DL (*deep learning*) mają już dość ugruntowaną renomę i w miarę precyzyjne definicje, chociaż różnice pomiędzy tymi pojęciami są dość subtelne. Ich charakterystykę można przedstawić m.in. poprzez dokonanie podziału na cel oraz wykorzystanie danych. Porównanie tych powiązanych pojęć przedstawione zostało w tabeli 3.1.

Sztuczna inteligencja okazała się doskonałym narzędziem do analizy i interpretacji istniejących danych, poprzez stosowanie podejścia określnego czasami jako dyskryminujące. Tymczasem generatywna sztuczna inteligencja ma niezwykłą zdolność do identyfikowania wzorców oraz tworzenia całkowicie nowej i oryginalnej treści od podstaw. Stanowi to podstawową różnicę między klasycznym (AI) i współczesnym (GenAI) podejściem. Aktualnie jest to możliwe dzięki dużym modelom językowym (LLM, *large language model*), potrafiącym przetwarzać ogromne zbiory informacji i generować odpowiedzi w języku naturalnym. Ponieważ LLM bazują na sztucznych sieciach neuronowych, GenAI w tabeli 3.1 sklasyfikowane zostało jako podzbiór DL. Teoretycznie jednak istnieje możliwość wykorzystania alternatywnych modeli ML. W takim przypadku klasyfikacja powinna ulec zmianie. Na razie konkurencyjne podejścia to przede wszystkim GAN (*generative adversarial networks*), VAE (*variational*

autoencoders) oraz modele dyfuzji (*diffusion models*). Wszystkie te metody wykorzystują różne typy sztucznych sieci neuronowych. Nie pojawiły się dotychczas żadne koncepcje mogące zagrozić ich dominacji.

W kontekście tematyki niniejszego rozdziału ważniejsze od niuansów technicznych, stojących za implementacją istniejących rozwiązań, są najpopularniejsze obszary zastosowań GenAI. Z pewnością należą do nich:

- generowanie obrazów,
- tworzenie i tłumaczenie tekstu,
- powiększanie zbioru danych o nowe, sztucznie stworzone obserwacje,
- odkrywanie związków chemicznych i leków,
- komponowanie muzyki,
- tworzenie środowisk i scenariuszy w grach komputerowych,
- monitorowanie i wykrywanie anomalii,
- interakcja z użytkownikami z czasie rzeczywistym,
- wsparcie przy pisaniu kodu.

Tabela 3.1. Charakterystyka porównawcza pojęć związanych z AI

Charakterystyka	AI	ML	DL	GenAI
Początki	wiek XX		wiek XXI	
	lata 50.	lata 80.	I dekada	II dekada
Powiązanie	–	podzbiór AI	podzbiór ML	podzbiór DL
Cel	rozwój systemów komputerowych wykonujących zadania, do których potrzebna jest ludzka inteligencja	dokonywanie przewidywań lub podejmowanie decyzji na podstawie istniejących danych	podejmowanie decyzji z użyciem sztucznych sieci neuronowych	generowanie nowych próbek danych, odzwierciedlających zależności występujące w zestawie danych treningowych
Podejście	symulacja ludzkiej inteligencji w celu rozwiązania złożonych problemów	wykorzystanie technik statystycznych do uczenia się na podstawie danych	wykorzystanie sztucznych sieci neuronowych z wieloma warstwami w celu wyodrębnienia cech i wzorców z danych	tworzenie modeli zdolnych do generowania nowej treści, przypominającej istniejące dane
Wykorzystanie danych	wykorzystanie różnych metod do naśladowania ludzkiej inteligencji w szerokim zakresie zastosowań	nauka na podstawie danych, aby tworzyć prognozy lub podejmować decyzje na podstawie nowych, niewidzianych danych	nauka bezpośrednio z danych, bez konieczności wskazywania zmiennych	generowanie nowych danych, które nie były częścią pierwotnego zestawu danych, ale mają podobne cechy

Źródło: opracowanie własne.

Lista ta nie stanowi zamkniętego zbioru stosowalności GenAI, gdyż każdego wręcz dnia powstają nowe, innowacyjne pomysły bazujące na generatywnych możliwościach AI.

3.2. GenAI – *status quo*

O popularności koncepcji wykorzystujących GenAI niech poświadczy kilka statystyk. Uruchomiony 30 listopada 2022 r. ChatGPT firmy OpenAI przyciągnął milion użytkowników w pięć dni. Dla porównania [Forsal.pl, 2023] Twitterowi zajęło to dwa lata, Facebookowi 10 miesięcy, a Netflixowi trzy i pół roku. Na przekroczenie granicy 100 mln użytkowników ChatGPT musiał czekać zaledwie dwa miesiące od swojej premiery [Wikipedia, 2023]. W 2024 r. już 92% firm z listy Fortune 500 używało produktu OpenAI [Binance Square, 2024]. Natomiast według badań ankietowych przeprowadzonych w 2023 r. [CCNEWS, 2023] aż 70% osób z pokolenia Z[1] wypróbowało narzędzia generatywnej sztucznej inteligencji. Przy tak dużej popularności programów wykorzystujących GenAI wśród osób fizycznych nie powinno dziwić zwiększone zainteresowanie przedsiębiorców. Potwierdzeniem tego faktu mogą być statystyki mówiące, że aż co piąty użytkownik generatywnej sztucznej inteligencji w Polsce korzysta z niej w celach zawodowych [Deloitte, 2023a]. Ciekawym zagadnieniem jest określenie obszarów, w których GenAI wykorzystywana jest w największym stopniu, a także jakiego rodzaju zastosowania przodują aktualnie wśród przedsiębiorstw.

W raporcie EY z maja 2024 r. [Ernst, Young LLP, 2024] podano szacunkowe wartości rzędu 30–40% wzrostu liczby zastosowań GenAI w przedsiębiorstwach w kilku początkowych miesiącach 2024 r. Stanowi to potwierdzenie przewidywań respondentów ankiety przeprowadzonej przez McKinsey, a mianowicie skokowy wzrost wykorzystania GenAI w przedsiębiorstwach z 33% w 2023 r. do 65% w 2024 r. [McKinsey, 2023]. Integracja z narzędziami wykorzystującymi generatywną AI nie jest jednak równomierna w przekroju przedsiębiorstw i sektorów gospodarki. Firmy o profilu technologicznym zdecydowanie przodują w zestawieniu. Według *AI in Business – Global Back to Contents Trends Report 2024* firmy Addepto [2024] aż 48% z nich przyznało się do stosowania GenAI w swoich produktach i usługach. Na dalszych miejscach znalazły się firmy z sektora handlu detalicznego i e-commerce, finansów i ubezpieczeń, opieki zdrowotnej (po 6,5%), a następnie produkcja przemysłowa, marketing i reklama, logistyka i transport, energetyka oraz edukacja (po 3%). Mimo

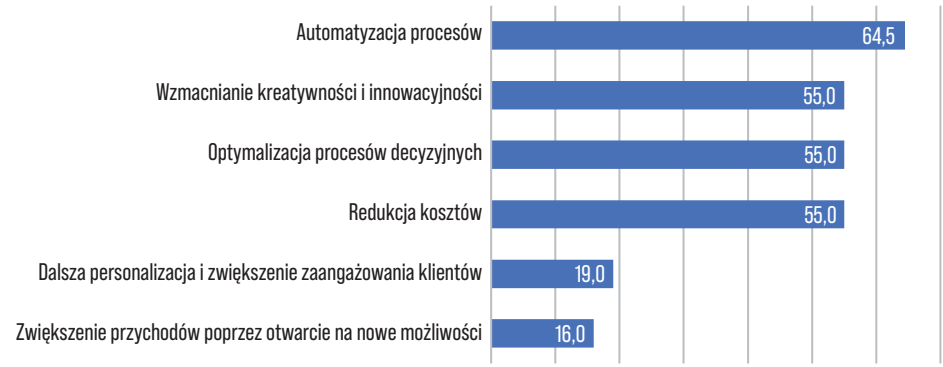
[1] *Generation Z* obejmuje osoby urodzone w latach 1997–2010 lub 1995–2012.

wyraźnego wzrostowego trendu te niewielkie procenty świadczą o tym, że natychmiastowe dostosowanie działalności firm do możliwości oferowanych przez GenAI nie jest możliwe i sporo czasu upłynie, zanim korzyści z wdrożenia rozwiązań opartych na generatywnej AI okażą się na tyle duże, aby dokonać jakościowego przełomu w polityce poszczególnych przedsiębiorstw. Należy również zwrócić uwagę, że sektor technologiczny niejako zmuszony jest do szybszego wdrożenia rozwiązań GenAI, przede wszystkim przez szybko rozwijającą się konkurencję. Zrozumiały jest jednak znacznie wolniejszy proces przyswajania GenAI w przypadku firm z innych sektorów gospodarki. Wiele z nich bowiem ograniczanych jest przez obowiązujące regulacje prawne (np. sektor medyczny czy finansowy), których zmiana czy dostosowanie wymaga dłuższego czasu. Dodatkowo nakład środków finansowych oraz wysiłek włożony w edukację pracowników, które są niezbędne przy tak gruntownych zmianach, stanowi często zbyt duże ryzyko, zwłaszcza w dynamicznie zmieniającej się technologicznej rzeczywistości. Nie można również pominąć jeszcze jednego aspektu wdrożeń nowych rozwiązań bazujących na generatywnej AI, a mianowicie pracowników zagrożonych utratą pracy. Bez wątpienia bowiem rewolucja GenAI będzie także miała niebagatelny wpływ na rynek pracy, zarówno eliminując, jak i tworząc wiele stanowisk. Z przewidywaniami dotyczącymi wpływu generatywnej sztucznej inteligencji na strukturę rynku pracy można zapoznać się m.in. poprzez lekturę artykułów Daco [2024], Rege'a i Hemachandrana [2024] czy Kalisha i Wolfa [2023]. Wykraczają one jednak poza zakres niniejszego opracowania.

Szybkość wdrażania GenAI w przedsiębiorstwach zależy bezpośrednio od tego, jak dużą wartość dodaną wnoszą proponowane rozwiązania. Warto zatem wymienić, jakie – spośród wielu dostępnych – zastosowania są często wykorzystywane w przedsiębiorstwach. Według przywoływanego już raportu Addepto [2024, s. 13] aż 61% respondentów wymieniło tworzenie ogólnie rozumianej treści, a 45% wskazało na zastosowanie NLP (*natural language processing*). Na kolejnych miejscach znalazły się rozszerzanie danych (*data augmentation*, 39%), wsparcie IT i rozwój oprogramowania (32%) oraz kreatywne projektowanie (29%). Większość z wymienionych zastosowań jest ściśle powiązana z sektorem IT, co nie powinno dziwić, skoro to przedsiębiorstwa o profilu technologicznym przodują we wdrożeniach GenAI w swojej działalności. Co więcej, wskazane zastosowania sprowadzają się do tych najpopularniejszych, rozpowszechnionych przez ChatGPT możliwościach stwarzanych przez GenAI. W głównej mierze bazują bowiem na zapytaniach, dających możliwość zaawansowanego wyszukiwania oraz generowania tekstu czy obrazów. Pozostałe, wskazane przez respondentów, zastosowania wiążą się z zadaniami z zakresu informatyki i *data science*. Tak ograniczony zakres wskazywanych zastosowań skłania do rozważań, czy wynika on z braku

pomysłów na przydatne w działalności przedsiębiorstw rozwiązania wykorzystujące możliwości GenAI, czy też może dłuższy proces wdrażania niż w przypadku dostępnych na rynku zastosowań w sektorze IT. Częściowej odpowiedzi na to pytanie mogą udzielić wyniki ankiety przeprowadzonej wśród tego samego grona respondentów (zestawiono je na rysunku 3.1).

Rysunek 3.1. Odsetek respondentów wskazujących na określone korzyści z generatywnej sztucznej inteligencji w biznesie (%)



Źródło: opracowanie własne na podstawie Addepto [2024].

Z danych przedstawionych na rysunku 3.1 wynika, że ponad połowa respondentów potrafiła wskazać aż cztery wymierne korzyści z generatywnej sztucznej inteligencji w biznesie. Duża część z nich w dalszym ciągu sprowadza się jednak do wykorzystania na potrzeby przedsiębiorstwa popularnych narzędzi GenAI, takich jak ChatGPT oferowany przez OpenAI czy Gemini firmy Google, jak również bardziej specjalistycznych Microsoft Copilot, Perplexity, Jasper czy Grammarly. Z pomocą tych narzędzi specjaliści do spraw sprzedaży mogą tworzyć szybciej i taniej artykuły oraz treści promocyjne, a programiści kod źródłowy programów. Nie ma jednak na razie innych spektakularnych i dostrzeganych przez większość uczestników rynku kierunków wykorzystania GenAI w bezpośredniej działalności przedsiębiorstw. Powodem mogą być m.in. wskazywane przez BCG [2024] niedobór talentów i umiejętności (62%), niejasne priorytety inwestycyjne (47%) i brak strategii na rzecz odpowiedzialnej sztucznej inteligencji (42%). Według tego samego raportu aż 90% przedsiębiorstw wybiera strategię oczekiwania, aż GenAI wyjdzie poza szum medialny, albo prowadzi eksperymenty z GenAI na małą skalę. Widać zatem mocny dysonans pomiędzy marketingowymi hasłami na temat innowacyjności oraz oczekiwaniami dotyczącymi potencjału generatywnej AI a faktycznie podejmowanymi działaniami ze strony firm. Z tej wstrzemięźliwości

w działaniu wybijają się firmy sektora technologicznego, dla których wspomniane antybodźce *de facto* nie istnieją lub przejawiają się w minimalnym stopniu. Nie zmienia to faktu, że generatywna sztuczna inteligencja pozostaje w kręgu ścisłego zainteresowania przedsiębiorstw nie tylko z sektora technologicznego.

3.3. GenAI w biznesie – potencjał i zagrożenia

Mimo deklaracji dotyczących chęci szybkiego wdrożenia rozwiązań opartych na GenAI większość przedsiębiorstw stosuje dostępne narzędzia w dość ograniczonym zakresie, a ewentualne dalsze inwestycje blokowane są przez szereg czynników, w tym brak wykwalifikowanych pracowników, wysokie koszty implementacji oraz duże i trudne do mitygacji ryzyko wdrożenia. Nie umniejsza to jednak dużego potencjału GenAI z punktu widzenia przedsiębiorstwa. Na potwierdzenie przytoczmy kilka statystyk.

W przeprowadzonym przez PWC [2024] badaniu wśród właścicieli firm 86% z nich wskazuje, że sztuczna inteligencja stała się powszechnie stosowana w ich firmach w codziennych operacjach biznesowych. Jednocześnie szacuje się, że sam ChatGPT pozwala na zwiększenie produktywności pracowników nawet o 40% [Talent Alpha, 2024]. Pokazuje to potencjalne kierunki transformacji dotychczasowych modeli pracy i potencjalny wzrost efektywności.

Poza zwiększaniem produktywności pracy oraz wymuszeniem zmiany profilu pracowników kluczową zaletą GenAI jest znacznie niższy próg wejścia dla nowych działań biznesowych oraz znaczne przyspieszenie badań i rozwoju poprzez projektowanie generatywne. Aktualnie widoczne jest to raczej na mocno specjalistycznych obszarach, takich jak sektor gamingowy [Limbasiya, 2024], muzyczny [Marr, 2023], [Moore, Acharya, 2023] czy farmaceutyczny [McKinsey, 2024]. W innych sektorach istnieje jednak również duży potencjał aplikacyjności, szczególnie na polu automatyzacji procesów i pracy ludzkiej w przedsiębiorstwach. Potencjał ten według badań Bain & Company [Singh, O'Keeffe, 2023] szacowny jest na co najmniej 40% w sektorach takich jak usługi administracyjne czy przemysł informacyjny (łącznie z telekomunikacją), ale tylko na niespełna 30% w sektorach usługowych.

Oprócz automatyzacji zadań jednym z częściej wymienianych zastosowań GenAI jest personalizacja doświadczenia klienta. To też jednak jeden z obszarów, w którym oczekiwania klientów zderzają się z preferowanym kierunkiem rozwoju firm zajmujących się obsługą klienta. Mimo rosnącej bowiem obecności chatbotów tylko 7% użytkowników deklaruje zaufanie do nich, w przeciwieństwie do aż 49% osób, które wykazują zaufanie do porad rzeczywistych ekspertów [Saldanha, Staehle, 2021].

Brak akceptacji klientów to tylko jeden z wielu problemów, napotykanych przez przedsiębiorstwa wdrażające GenAI [University of Leed, 2024]. Do tych częściej wymienianych należy z pewnością ograniczony dostęp do najnowszych technologii. Tylko nieliczne firmy mogą udźwignąć wysokie koszty początkowe, często przekraczające milion USD, za infrastrukturę informatyczną, wydatki na szkolenia oraz dostęp do danych. W przeciwnym razie są uzależnione od potentatów na rynku informatycznym, oferujących dostęp do usług w zakresie GenAI [Wu, 2024].

Brak zaufania do narzędzi wykorzystujących GenAI ma swoje uzasadnienie w często kwestionowanej ich wiarygodności, która manifestuje się generowaniem niedokładnych, niezweryfikowanych lub po prostu nieprawdziwych informacji. Zjawisko to znane jest pod nazwą halucynacji AI [IBM, 2023]. Przedsiębiorstwa nie mogą sobie natomiast pozwolić na uszczerbek w reputacji związany z uzyskaniem lub przekazaniem błędnych danych. W szczególności groźne są wszelkie odpowiedzi dyskryminujące określone grupy demograficzne bądź osoby ze względu na pochodzenie, płeć czy religię. Stąd decyzje o braku rozpoczęcia lub wstrzymaniu prac wdrożeniowych.

Jednym z największych zagrożeń dla przedsiębiorstw z punktu widzenia wdrożenia GenAI są obawy prawne. Dyskusyjną kwestią pozostają prawa własności intelektualnej, które obejmują naruszenie praw autorskich i znaków towarowych [Vadapalli, 2023]. Dzieła licznych twórców i artystów, dostępne w Internecie, są bowiem przez wiele programów wykorzystywane w procesie trenowania modeli AI. Rodzi to poważne obawy dotyczące oryginalności treści generowanych przez te systemy oraz obejmujących je praw autorskich, a tym samym stwarza realne ryzyko dla działalności firmy. Wyrazem dezaprobaty dla takiego zachowania potentatów technologicznych są pozwy niezadowolonych autorów i redakcji, np. pozew „The New York Times" przeciwko OpenAI i Microsoft [Rotkiewicz, 2023], w którym podnoszone są argumenty wykorzystania bez zgody milionów artykułów do treningu modeli językowych takich jak ChatGPT.

W ostatnich latach dużą wagę przykłada się do ekologizacji, w tym tzw. śladu węglowego (*carbon footprint*) pozostawianego w wyniku działalności człowieka. Ilość energii elektrycznej niezbędnej do trenowania modeli GenAI jest zazwyczaj liczona w dziesiątkach tysięcy megawatogodzin i stale rośnie [de Vries, 2023]. Przekłada się to na tysiące ton metrycznych emisji dwutlenku węgla (w zależności m.in. od lokalizacji centrów obliczeniowych). Dla porównania międzykontynentalny lot w obie strony to około jednej tony wyemitowanego CO_2. Wspomagana przez GenAI odpowiedź na zapytanie pozostawia nawet dziesięciokrotnie większy ślad węglowy niż klasyczne zapytanie do wyszukiwarki internetowej. Ponieważ istnieją klienci, którzy zwracają uwagę na nastawienie przedsiębiorstw do ochrony środowiska, jak również regulacje

prawne w niektórych państwach narzucają konieczność ograniczenia negatywnego wpływu na środowisko, kwestie ekologii muszą być brane pod uwagę przez przedsiębiorstwa przy podejmowaniu decyzji związanych z inwestowaniem w rozwiązania wykorzystujące GenAI.

Ostatni z poruszanych w niniejszym rozdziale aspektów związanych z GenAI z punktu widzenia przedsiębiorstw stanowią dylematy natury etycznej. Jak wiele technologii, tak również GenAI może być wykorzystane do monetyzacji niezgodnej z przyjętymi standardami moralnymi. Przykładem takich działań są z pewnością dezinformacja (*fake news*), trolling oraz tzw. *deepfake*, czyli obróbka obrazu w celu stworzenia nieprawdziwych sytuacji z wykorzystaniem realnych lub wygenerowanych wirtualnie osób. Spektakularnym i zarazem przerażającym przykładem jest *deepfake* w branży pornograficznej [Wiggers, 2023]. Realistyczne zdjęcia i wideo z możliwością podstawienia twarzy dowolnej osoby, bez jej wiedzy i zgody, może prowadzić do skompromitowania nie tylko celebrytów, ale i pracowników przedsiębiorstw, a tym samym wpłynąć na reputację firmy. Można wyobrazić sobie scenariusze, w których wpływa się w ten sposób na wyniki finansowe przedsiębiorstw, a nawet wyniki wyborów na poziomie lokalnym czy ogólnopaństwowym.

Podsumowanie

W niniejszym rozdziale poruszono wybrane, istotne z punktu widzenia przedsiębiorstw, aspekty wykorzystania generatywnej sztucznej inteligencji, jej potencjału rozwojowego oraz zagrożeń związanych z dynamicznym rozwojem technologicznym i chęcią integracji rozwiązań z aktualnie obowiązującymi procesami w przedsiębiorstwach. Część z nich opiera się na przewidywaniach wydedukowanych na podstawie badań ankietowych i w związku z tym podane statystyki mogą mocno rozmijać się z faktycznym rozwojem tego obszaru nauki i biznesu. Należy też wziąć poprawkę na kwestie promocji, wpływające na wyolbrzymianie niektórych osiągnięć ogłaszanych przez firmy. Warto także zwrócić uwagę na znaczącą różnicę pomiędzy rzeczami, które można zrealizować pod względem technologicznym, a takimi, których wdrożenie jest opłacalne z punktu widzenia przedsiębiorstwa. Tym samym wdrażanie GenAI w przedsiębiorstwach można traktować jako swego rodzaju problem decyzyjny, który w porównaniu z erą przed generatywną AI, jest bardziej złożony i obarczony większą niepewnością.

Wymienione aspekty nie wyczerpują w żadnym przypadku tematyki GenAI. Istnieją znane kwestie, które bezpośrednio lub pośrednio wpływają na działalność przedsiębiorstwa, a są konsekwencją generatywnej AI. Wiele aspektów jednak zostanie dopiero zidentyfikowanych wraz ze wzrostem penetracji rynku przez narzędzia bazujące na GenAI. Również przedstawione opisy i przykłady należy traktować raczej jako sygnalizację istnienia konkretnego problemu, a nie szczegółowy opis jego rozwiązania. Z pewnością jednak generatywna AI ma potencjał na zrewolucjonizowanie rynku pracy i działalności przedsiębiorstw. Czy to nastąpi, dowiemy się w najbliższych latach.

Bibliografia

Adobe (2023). *Co to jest i jak działa generatywna sztuczna inteligencja?*, https://www.adobe.com/pl/products/firefly/discover/how-generative-ai-work.html (dostęp: 28.07.2024).

Addepto (2024). *AI in Business – Global Back to Contents Trends Report 2024*, https://addepto.com/wp-content/uploads/2024/04/GenAI-in-Business_Global-Trends-Report-2024.pdf (dostęp: 28.07.2024).

Apotheker, J., Duranton, S., Lukic, V., de Bellefonds, N., Iyer, S., Bouffault, O., de Laubier, R. (2024). *From Potential to Profit with GenAI*, https://www.bcg.com/publications/2024/from-potential-to-profit-with-genai (dostęp: 28.07.2024).

Binance Square (2024). *Sam Altman Pushes ChatGPT Mass Adoption Among Fortune 500 Companies: Report*, https://www.binance.com/pl/square/post/6676169359089?ref=516042411 (dostęp: 28.07.2024).

CCNEWS (2023). *Pokolenie Z zdecyduje o przyszłości GenAI*, https://ccnews.pl/2023/10/11/pokolenie-z-zdecyduje-o-przyszlosci-genai/ (dostęp: 28.07.2024).

Deloitte (2023a). *Co piąty użytkownik generatywnej sztucznej inteligencji w Polsce korzysta z niej w celach zawodowych*, https://www2.deloitte.com/pl/pl/pages/press-releases/articles/Co-piaty-uzytkownik-generatywnej-sztucznej-inteligencji-w-Polsce-korzysta-z-niej-w-celach-zawodowych.html (dostęp: 28.07.2024).

Deloitte (2023b). *Czym jest generatywna sztuczna inteligencja? Podstawowe zagadnienia dziedziny GenAI*, https://www2.deloitte.com/pl/pl/pages/risk/articles/Czym-jest-generatywna-sztuczna-inteligencja.html (dostęp: 28.07.2024).

Daco, G. (2024). *How GenAI Will Impact the Labor Market*, https://www.ey.com/en_gl/insights/ai/how-gen-ai-will-impact-the-labor-market (dostęp: 28.07.2024).

Ernst & Young LLP (2024). *Is Generative AI Beginning to Deliver on Its Promise in India?*. AIdea of India update. India.

Forsal.pl (2023). *ChatGPT przyciągnął milion użytkowników w 5 dni. Twitterowi zajęło to dwa lata*, https://forsal.pl/lifestyle/technologie/artykuly/8645576, chatgpt-jak-szybko-przyciaga-uzytkownikow.html (dostęp: 28.07.2024).

IBM (2023). *What Are AI Hallucinations?*, https://www.ibm.com/topics/ai-hallucinations (dostęp: 28.07.2024).

Kalish, I., Wolf, M. (2023). *Generative AI and the Labor Market: A Case for Techno-Optimism*, https://www2.deloitte.com/us/en/insights/economy/generative-ai-impact-on-jobs.html (dostęp: 28.07.2024).

Limbasiya, J. (2024). *Game on: The Evolution of Gaming Through Generative AI Innovation*, https://www.cio.com/article/1294946/game-on-the-evolution-of-gaming-through-generative-ai-innovation.html (dostęp: 28.07.2024).

Lorenz, P., Perset, K., Berryhill, J. (2023). Initial Policy Considerations for Generative Artificial Intelligence, *OECD Artificial Intelligence Papers*, September 18. DOI: 10.1787/fae2d1e6-en.

Marr, B. (2023). *Generative AI Is Revolutionizing Music: The Vision for Democratizing Creation*, https://www.forbes.com/sites/bernardmarr/2023/10/05/generative-ai-is-revolutionizing-music-loudlys-vision-for-democratizing-creation/ (dostęp: 28.07.2024).

McKinsey (2023). *The State of AI: How Organizations Are Rewiring to Capture Value*, https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai (dostęp: 28.07.2024).

McKinsey (2024). *Generative AI in the Pharmaceutical Industry: Moving from Hype to Reality*, https://www.mckinsey.com/industries/life-sciences/our-insights/generative-ai-in-the-pharmaceutical-industry-moving-from-hype-to-reality (dostęp: 28.07.2024).

Moore, J., Acharya, A. (2023). *The Future of Music: How Generative AI Is Transforming the Music Industry*, https://a16z.com/the-future-of-music-how-generative-ai-is-transforming-the-music-industry/ (dostęp: 28.07.2024).

PWC (2024). *AI Business Predictions*, https://www.pwc.com/us/en/tech-effect/ai-analytics/ai-business-survey.html (dostęp: 28.07.2024).

Rege, M., Hemachandran, K. (2024). *Tommie Expert: Generative AI's Real-World Impact on Job Markets*, https://news.stthomas.edu/generative-ais-real-world-impact-on-job-markets/ (dostęp: 28.07.2024).

Rotkiewicz, M. (2023). *„New York Times" pozywa OpenAI i Microsoft. Proces byłby bez precedensu*, https://www.polityka.pl/tygodnikpolityka/nauka/2240193,1, new-york-times-pozywa-openai-i-microsoft-proces-bylby-bez-precedensu.read (dostęp: 28.07.2024).

Saldanha, K., Staehle, T. (2021). *Guide Insurance Customers to Safety and Well-Being*, https://www.accenture.com/us-en/insights/insurance/guide-insurance-customers-safety-well-being (dostęp: 28.07.2024).

SAP (2023). *Co to jest generatywna sztuczna inteligencja?*, https://www.sap.com/poland/products/artificial-intelligence/what-is-generative-ai.html (dostęp: 28.07.2024).

Singh, R., O'Keeffe, D. (2023). *How Generative AI Will Supercharge Productivity, Bain and Company*, https://www.bain.com/insights/how-generative-ai-will-supercharge-productivity-snapchart/ (dostęp: 28.07.2024).

Talent Alpha (2024). *The World of Work AI-ed. The Future of Work Report*, https://talent-alpha.com/future-of-work-report-2024-download/ (dostęp: 28.07.2024).

University of Leeds (2024). *Strengths and Weaknesses of Gen AI*, https://generative-ai.leeds.ac.uk/intro-gen-ai/strengths-and-weaknesses/ (dostęp: 28.07.2024).

Vadapalli, P. (2023). *The Pros and Cons of Generative AI*, https://www.upgrad.com/blog/pros-and-cons-of-generative-ai/ (dostęp: 28.07.2024).

Wiggers, K. (2023). *As AI Porn Generators get Better, the Stakes get Higher*, https://techcrunch.com/2023/07/21/as-ai-porn-generators-get-better-the-stakes-raise (dostęp: 28.07.2024).
Vries de, A. (2023). The Growing Energy Footprint of Artificial Intelligence, *Joule*, 7(10), s. 2191–2194. DOI: 10.1016/j.joule.2023.09.004.
Wikipedia (2023). *ChatGPT*, https://pl.wikipedia.org/wiki/ChatGPT (dostęp: 28.07.2024).
Wu, M. (2024). *The GenAI Economic Engine Will Sputter and Stall Without an Inclusive Business Model*, https://www.fastcompany.com/91166266/the-genai-economic-engine-will-sputter-and-stall-without-an-inclusive-business-model (dostęp: 28.07.2024).

4.0

ZARZĄDZANIE INNOWACJAMI WE WDRAŻANIU SZTUCZNEJ INTELIGENCJI W ORGANIZACJI

Ryszard Ćwiertniak
Uniwersytet Ekonomiczny w Krakowie

Streszczenie

W rozdziale omówiono rolę zarządzania innowacjami w trakcie wdrażania sztucznej inteligencji (*artificial intelligence*, AI) w organizacji. W treści przedstawiono, w jaki sposób zastosowanie AI wpływa na automatyzację procesów oraz poprawę efektywności operacyjnej. W dynamicznie zmieniającym się otoczeniu biznesowym interdyscyplinarne podejście do działalności innowacyjnej oraz integracja nowych technologii z procesami tworzenia produktów i usług mają kluczowe znaczenie dla utrzymania konkurencyjności. Zwrócono uwagę na aspekty związane z ryzykiem, które należy uwzględnić podczas implementacji nowych rozwiązań technologicznych. Podkreślono także znaczenie odpowiedniego przygotowania pracowników, dojrzałości kultury organizacyjnej oraz zarządzania zmianą, aby zapewnić zgodność rozwiązań z regulacjami oraz akceptację użytkowników końcowych. Dodatkowo, przedstawiono matrycę pomiaru, która finalizuje rozważania na temat oceny zdolności organizacji do implementacji sztucznej inteligencji, przekładając na praktyczne działania ważne aspekty zarządzania innowacjami, takie jak integracja technologii, gotowość do zmian i efektywność operacyjna.

Słowa kluczowe: sztuczna inteligencja, integracja technologii, wdrażanie sztucznej inteligencji, innowacje biznesowe AI

Wprowadzenie

Sztuczna inteligencja (*artificial intelligence*, AI) posiada potencjał do tworzenia przełomowych innowacji [Grashof, Kopka, 2023]. Z kolei zarządzanie innowacjami jest obecnie ważniejsze niż kiedykolwiek wcześniej, stanowi istotny czynnik rozwoju organizacji. W rozdziale tym przyjęto, że interdyscyplinarne i międzyfunkcyjne[1] podejście do innowacji AI jest kluczowe w złożonym otoczeniu biznesowym. Organizacje, aby nadążyć za trendami, integrują nowe technologie z procesem tworzenia produktów lub usług. Jednocześnie powinny one zachować daleko idącą ostrożność.

Według autora zastosowanie narzędzi sztucznej inteligencji w organizacjach powinno być postrzegane jako oddzielna dziedzina – działalność innowacyjna[2]. Automatyzacja rutynowych zadań oraz wsparcie w zaawansowanych analizach danych pozwalają odkrywać nowe możliwości, takie jak: optymalizacja procesów, zwiększenie wydajności, lepsze zrozumienie rynku, szybsze podejmowanie decyzji oraz tworzenie innowacyjnych produktów i usług, które lepiej odpowiadają na potrzeby klientów. W tym kontekście trzeba jednak pamiętać, że „innowacja jest to wszelka, z założenia korzystna zmiana w różnych obszarach działalności organizacji, wnosząca postęp w stosunku do stanu istniejącego. Często ma ona charakter ewolucyjnego poprawiania rzeczy istniejących, ocenianego pozytywnie w świetle kryteriów danej organizacji" [Kozioł, 2013, s. 69–70]. Należy też stwierdzić, że zastosowanie AI w organizacji obejmuje złożony proces, tj. poszukiwanie wiedzy i perswazję, podejmowanie decyzji, wdrożenie oraz potwierdzenie zmiany [Carayannis i in., 2024, s. 17–21].

Jednocześnie, z rosnącą dostępnością nowych technologii, konkurencja może minimalizować przewagę innowatorów poprzez stosowanie podobnych narzędzi. W związku z tym organizacje powinny stale doskonalić swoje działania, inwestować w nowe strategie, budować relacje z klientami oraz rozwijać kompetencje swoich pracowników.

Celem tego rozdziału jest zatem charakterystyka wybranego etapu zarządzania innowacjami opartymi na AI – dyfuzji innowacji biznesowych, dlatego w treści omówiono kluczowe czynniki, które mogą wpływać na przyswajanie nowych narzędzi AI w organizacji. W dalszej części wykazano, że efektywne zarządzanie innowacjami w obszarze wdrażania AI wymaga strategicznego podejścia, obejmującego zarówno aspekty technologiczne, jak i organizacyjne [Ćwiertniak, 2024, s. 12–13], ponadto:

---

1 Zespoły interdyscyplinarne lub międzyfunkcyjne składają się z osób z różnych obszarów specjalizacji (dyscyplin) albo działów (funkcji) pracujących razem nad wspólnym celem w organizacji.

2 Działania rozwojowe, finansowe i komercyjne podejmowane przez organizację w celu tworzenia innowacji.

- integracja AI z istniejącymi systemami informatycznymi oraz procesami biznesowymi jest niezbędna do pełnego wykorzystania potencjału nowej technologii;
- sukces aplikacji AI zależy w dużej mierze od dojrzałości organizacyjnej oraz kompetencji pracowników.

Zakres badań literaturowych został ograniczony do obszaru organizacji, ze szczególnym uwzględnieniem innowacyjności, na którą duży wpływ mają pracownicy. W przeprowadzonych analizach szczególną uwagę zwrócono na znaczenie dojrzałości kulturowej. Dojrzała kultura organizacyjna, otwarta na innowacje, elastyczna i wspierająca współpracę między zespołami jest uznawana za ważny czynnik sukcesu. W rozdziale przedstawiono także propozycję matrycy oceny zdolności organizacji do wdrażania sztucznej inteligencji (OZO), która mierzy ważne dziedziny zarządzania innowacjami, takie jak funkcjonalność, wykonalność oraz opłacalność projektu innowacyjnego opartego na systemach AI.

Współczesne modele sztucznej inteligencji[3] mogą zmieniać wybrane funkcje biznesowe[4], dlatego międzyfunkcyjne zarządzanie innowacjami[5] powinno zapewnić skuteczne wdrożenie AI poprzez poprawę zaangażowania pracowników. Pomaga ono uwzględnić ryzyko etyczne, prawne i bezpieczeństwa, łącząc wiedzę z takich dziedzin jak prawo, IT i zarządzanie. Jasno ustalone role użytkowników końcowych[6] promują bowiem kulturę odpowiedzialności, a współpraca między działami w strukturze organizacyjnej ułatwia wymianę wiedzy i dobrych praktyk. Taki sposób zarządzania oznacza przejrzystość, odpowiedzialność oraz celowe wykorzystanie technologii [Sinha i in., 2024].

3 Sztuczna inteligencja to dziedzina informatyki zajmująca się tworzeniem systemów i programów, które naśladują ludzkie myślenie oraz zachowanie, umożliwia maszynom wykonywanie zadań wymagających inteligencji, takich jak rozpoznawanie obrazów, przetwarzanie języka naturalnego czy podejmowanie decyzji. Z kolei model AI to matematyczna reprezentacja procesów myślowych, wykorzystywana przez algorytmy do rozwiązywania określonych problemów na podstawie danych treningowych. Dodatkowo innowacyjny model sztucznej inteligencji to zaawansowana wersja standardowego modelu, która wprowadza nowatorskie rozwiązania, takie jak większa precyzja w przewidywaniu, lepsza analiza wyników czy zdolność do adaptacji w dynamicznie zmieniającym się środowisku, przynosi organizacjom przewagę konkurencyjną i realizację bardziej złożonych działań.

4 Funkcje biznesowe – zbiór zadań, które organizacja powinna regularnie wykonywać w celu komercjalizacji produktu lub usługi.

5 Międzyfunkcyjne zarządzanie to współpraca pomiędzy różnymi działami organizacji, takimi jak IT, marketing czy produkcja, w celu skutecznego wdrażania technologii, typu sztuczna inteligencja. Dzięki temu różne zespoły wspólnie pracują nad realizacją celów firmy, pozwalając na poprawę integracji nowych technologii z codziennymi działaniami firmy.

6 Użytkownik końcowy to osoba, która ostatecznie korzysta z produktu, usługi lub oprogramowania. W kontekście biznesowym i technologicznym użytkownik końcowy jest odbiorcą, dla którego produkt został opracowany.

Organizacje, które integrują nowe narzędzia IT ze swoimi ogólnymi strategiami, mogą nie tylko wykorzystać pełny potencjał sztucznej inteligencji, ale także przyczynić się do poprawy efektywności operacyjnej oraz tworzenia wartości dla klientów. To z kolei może wzmocnić ich pozycję na rynku.

4.1. Strategiczne zarządzanie innowacjami we wdrażaniu AI – standaryzacja i integracja technologii

Generatywna sztuczna inteligencja (GenAI) znajduje się obecnie na wczesnym etapie rozwoju technologii oraz dyfuzji innowacji[7], pomimo to nowe narzędzia zaczynają przejmować wybrane funkcje biznesowe, wcześniej już zautomatyzowane, np. ERP (*enterprise resource planning*), CRM (*customer relationship management*) itd.

Kompleksowa standaryzacja odnosi się do synergii procesów technologicznych i organizacyjnych, co pozwala na integrację innowacyjnych rozwiązań AI w strukturach organizacji, zapewniając tym samym spójność i efektywność operacyjną. Nowa technologia może umożliwić małym organizacjom efektywne konkurowanie z większymi firmami, jednak wymaga to odpowiednich zasobów i strategii wykorzystania zaawansowanych narzędzi, obniżając barierę ich wejścia na rynek. Ważnym czynnikiem w tym procesie jest organizacyjna dojrzałość, którą mogą osiągnąć nawet mniejsze organizacje. Dzięki zwinnej strukturze mniejsze firmy adaptują się do nowych technologii oraz podejmują decyzje szybciej niż duże korporacje. To pozwala im elastycznie reagować na nowe trendy i innowacje, integrując skutecznie narzędzia AI w działalności operacyjne [Miric, 2023; Porter i in., 2015].

Z kolei normalizacja w organizacji oznacza działania zgodnie z ustalonymi procedurami. Przynosi ona korzyści w postaci spójności, efektywności i wysokiej jakości realizacji zadań, a także ułatwia porównywanie wyników i procesów z innymi organizacjami. Może jednak wiązać się również z negatywnymi skutkami, takimi jak biurokratyzacja procesu decyzyjnego, która prowadzi do spowolnienia reakcji organizacji na dynamicznie zmieniające się otoczenie rynkowe. Zbyt formalne procesy mogą utrudniać dopasowanie się do nowych wyzwań i ograniczać elastyczność działania. Ujednolicenie to może obejmować różne dziedziny zarządzania, w tym zarządzanie jakością, procesami, projektami oraz innowacjami i technologiami.

Według teorii innowacji Josepha Schumpetera innowatorzy to osoby, które wprowadzają na rynek przełomowe produkty, technologie lub usługi, przekształcając tym

[7] Innowatorzy dopiero niedawno zaczęli informować organizacje o korzyściach stosowania generatywnej sztucznej inteligencji [zob. Ćwiertniak, 2024, s. 39].

samym całe rynki za pomocą technologii tzw. twórczej destrukcji. Imitatorzy natomiast adaptują te nowości, nie wprowadzają przełomowych zmian, ale usprawniają istniejące rozwiązania i rozprzestrzeniają je na szeroką skalę. Granica między nimi polega na tym, że innowatorzy kreują możliwości, podczas gdy imitatorzy optymalizują i rozwijają istniejące pomysły [Ćwiertniak, 2024, s. 14]. Powszechny dostęp do nowych narzędzi technologicznych oznacza, że imitatorzy mogą korzystać z podobnych systemów. Mimo że ułatwia to tworzenie nowych produktów, jednocześnie pozwala konkurentom na kopiowanie pomysłów i produktów. W takim środowisku, w zależności od branży, sukces może bardziej zależeć od skutecznego zarządzania i relacji z klientami niż od samego postępu technicznego, choć w niektórych obszarach postęp ten może odgrywać ważną rolę, np.:

- nowe systemy wzmacniają pozycję dostawców technologii, ponieważ firmy budujące swoje produkty na bazie oferowanych rozwiązań muszą akceptować warunki ustalane przez producentów systemów, ograniczając ich swobodę działania;
- nowe platformy IT mogą sprzyjać tworzeniu podobnych, mniej innowacyjnych produktów. Chociaż obniżają one koszty oraz bariery wejścia na rynek, jednocześnie ograniczają unikatowość produktów opracowywanych z ich wykorzystaniem. To zjawisko może ograniczać różnorodność i innowacyjność produktów i usług na rynku[8].

Międzyfunkcyjne podejście do projektów AI polega na integracji wiedzy i umiejętności z różnych działów organizacji w celu skutecznego zarządzania sztuczną inteligencją. Zakłada ono współpracę między różnymi zespołami, takimi jak IT, marketing, produkcja, HR, prawny i inne, aby maksymalnie wykorzystać potencjał nowej technologii. Zalety standaryzowanej organizacji obejmują spójność i przewidywalność działań dzięki ujednoliconym procedurom. Uproszczenie i optymalizacja procesów prowadzą do zwiększenia efektywności operacyjnej, co może przekładać się na redukcję kosztów i poprawę konkurencyjności w branży oraz na rynku.

4.2. Integracja AI z systemami informatycznymi i procesami biznesowymi

Integracja narzędzi AI z istniejącymi systemami informatycznymi pozwala na efektywne wykorzystanie dostępnych zasobów i danych, co jest kluczowe dla uzyskania korzyści zarówno dla organizacji, jak i użytkowników końcowych. Niemniej wprowadzenie nowych technologii wiąże się także z wyzwaniami. Organizacje powinny być

[8] Znormalizowana technologia zwiększa mobilność pracowników na rynku pracy, ponieważ nie muszą oni uczyć się nowych programów od podstaw przy zmianie zatrudnienia.

gotowe na zakłócenie modeli biznesu i na tyle zwinne, aby adaptować się do nowych warunków [Nagarajan, 2023].

Liczba organizacji korzystających z botów generatywnej AI rośnie dynamicznie[9], zdecydowanie szybciej niż świadomość pracowników o potencjalnych zagrożeniach związanych z wykorzystaniem nowych narzędzi w praktyce. Jak wynika z najnowszych badań *Cyberhaven Labs*, ilość danych wprowadzanych do systemów AI[10] przez pracowników tylko w okresie od marca 2023 r. do marca 2024 r. wzrosła niemal sześciokrotnie [2024].

LLMs (duże modele językowe) generują prognozy przyszłych wzorców na podstawie analizy danych historycznych, jednak nie przewidują przyszłości w dosłownym sensie, a raczej identyfikują prawdopodobne wyniki na podstawie wzorców z przeszłości. Są to zaawansowane modele sztucznej inteligencji, które zostały przeszkolone na dużych zbiorach danych tekstowych w celu rozumienia, generowania i przetwarzania ludzkiego języka. Wykorzystują one techniki głębokiego uczenia, a w szczególności sieci neuronowe, aby przewidywać kolejne słowa lub frazy na podstawie wcześniejszego kontekstu. W przeciwieństwie do modeli prognostycznych, typu regresja liniowa, LLMs są zdolne do bardziej złożonych zadań, takich jak generowanie tekstu, tłumaczenie, odpowiadanie na pytania czy prowadzenie konwersacji. Modele te nie znają znaczenia pojęć „prawda" czy „fałsz", ponieważ są trenowane na danych internetowych i dostarczają odpowiedzi, które są statystycznie prawdopodobne w publicznych repozytoriach [Desai i in., 2023].

Negatywne skutki używania algorytmów w procesach wewnętrznych organizacji obejmują kilka ważnych kwestii. Po pierwsze, sztuczna inteligencja może pomóc pracownikom wykonującym powtarzalne zadania, automatyzując ich procesy i zwiększając wydajność. W przypadku kluczowych pracowników zbyt duże poleganie na AI może jednak obniżyć ich motywację, innowacyjność i zaangażowanie oraz prowadzić do spadku satysfakcji i trudności z utrzymaniem talentów, zwłaszcza jeśli ograniczo-

9 Na przykład w strukturze organizacyjnej korporacji IT – Nvidia pracuje obecnie 65 botów (*generative AI „agents"*).

10 System sztucznej inteligencji można zdefiniować jako oprogramowanie lub system komputerowy, który wykorzystuje techniki i algorytmy do realizacji określonych zadań. Definicja ta może obejmować zarówno systemy, które wykorzystują komponenty AI jako jedną z wielu technologii, jak i te, w których AI jest kluczowym elementem, napędzającym całość funkcjonowania. Systemy, w których AI jest „głównym silnikiem", to takie, gdzie procesy decyzyjne, automatyzacja lub analizy oparte są przede wszystkim na algorytmach uczenia maszynowego, sieciach neuronowych lub innych technikach AI. Z kolei systemy, które jedynie zawierają komponenty AI, mogą wykorzystywać tę technologię do wsparcia lub automatyzacji wybranych zadań, ale nie są całkowicie zależne od AI w swoim funkcjonowaniu. Przykładami mogą być zarówno autonomiczne systemy zarządzania produkcją, w których AI jest kluczowym elementem optymalizującym procesy, jak i systemy CRM, które korzystają z AI do analizy danych klientów, ale funkcjonują także z innymi technologiami.

na zostanie ich autonomia i możliwości twórcze. Po drugie, skuteczność modelu LLM dotyczy często samodzielnych zadań, na których model już został przeszkolony, zniechęcając kluczowych pracowników do kreatywnych działań i poszukiwania innowacyjnych rozwiązań [Duarte, 2024; Waber, Fast, 2024].

W tradycyjnych systemach sztucznej inteligencji proces podejmowania decyzji często pozostaje niejasny dla użytkowników, co jest efektem tzw. czarnej skrzynki. Właśnie dlatego stworzono wyjaśniające systemy AI, które pozwalają zrozumieć, jak narzędzie rozwiązuje złożone problemy decyzyjne. Modele tłumaczące algorytmy IT budują zaufanie wśród pracowników poprzez zapewnienie przejrzystości działania systemu, sprawdzenie zabezpieczeń oraz błędów. Dzięki temu użytkownicy mogą mieć większą pewność, że system działa zgodnie z ich oczekiwaniami. Ponadto zrozumienie funkcjonowania modeli AI umożliwia ich optymalizację, co prowadzi do sprawnych decyzji opartych na analizie danych biznesowych [Ćwiertniak, 2024, s. 44–48]. Korzyści te mogą się jednak różnić w zależności od wybranego przypadku oraz jakości baz danych [PwC, 2018].

4.3. Dojrzałość kultury organizacyjnej

We wdrażaniu sztucznej inteligencji w organizacji istotnym pojęciem jest autonomia pracowników, zwłaszcza ekspertów oraz specjalistów, których kompetencje są wprowadzane do bazy danych AI. Chodzi o to, że im więcej procesów i decyzji zostaje zautomatyzowanych, tym bardziej może zostać ograniczona samodzielność tych pracowników w podejmowaniu decyzji czy rozwiązywaniu problemów. Zastosowanie AI może zmniejszyć ich wpływ na kształtowanie rozwiązań, zaangażowanie oraz poczucie kontroli nad własną pracą.

Dojrzałość kultury organizacyjnej to poziom przygotowania oraz zdolność organizacji do przyjęcia innowacyjnych zmian. Obejmuje gotowość pracowników do nauki, elastyczność w podejściu do innowacji oraz współpracę między zespołami. Dojrzała organizacja jest otwarta na nowe rozwiązania, toleruje popełnianie błędów oraz odznacza się zdolnością do zarządzania zmianą.

Dlatego główny nurt dyskusji związanej z zastosowaniem AI dotyczy stwierdzenia – nowe narzędzia AI nie będą skuteczne w organizacji, jeżeli nie zaakceptują ich użytkownicy końcowi [Kelly i in., 2023]. Aby przekonać pracowników do nowości, liderzy zmian powinni odkryć genezę trzech konfliktów interesów. Po pierwsze, użytkownicy mogą nie dostrzegać korzyści dla siebie. Po drugie, mogą być obciążeni dodatkowymi obowiązkami w trakcie aplikacji systemów. Po trzecie, mogą stracić niezależność.

Menedżerowie powinni więc nie tylko przewidzieć trudności i nimi zarządzać, ale także budować bazę interdyscyplinarnej wiedzy poprzez odpowiednie angażowanie się w nowe technologie. W przeciwnym razie wartość nowych rozwiązań może okazać się niezadowalająca i w efekcie wpłynąć negatywnie na poziom satysfakcji pracowników [BCG, 2024; Kellogg i in., 2023, s. 40–48; McKinsey, Company, 2023].

W artykule pt. *Five Myths About Generative AI That Leaders Should Know*, opublikowanym przez Wharton School of the University of Pennsylvania, autorzy omawiają pięć powszechnych mitów związanych z generatywną sztuczną inteligencją, które mogą wpływać na jej postrzeganie oraz zastosowanie. Celem niniejszej publikacji jest potrzeba szkolenia liderów, aby wyzbyli się tych przekonań, które często towarzyszą rosnącej popularności AI, oraz wskazanie, jaki mogą mieć wpływ na podejmowanie decyzji. Snyder i Velasterui [2024] szczegółowo analizują przesadne oczekiwania związane z nową technologią, aby pomóc menedżerom w lepszym zrozumieniu potencjału nowej technologii.

- **Mit 1: Narzędzia generatywnej AI są dostępne za darmo lub niewiele kosztują.**
  Pomimo że niektóre narzędzia AI mogą być dostępne za darmo lub w niskiej cenie, profesjonalne rozwiązania często wymagają dużej inwestycji. Koszty te obejmują budowę infrastruktury, wsparcie techniczne oraz dostosowanie do specyficznych potrzeb organizacji.
- **Mit 2: AI zawsze generuje wysokiej jakości treści.**
  AI nie zawsze tworzy poprawne treści, ponieważ jej działanie opiera się na danych, na których została wytrenowana. Jakość tych danych ma kluczowe znaczenie, a bez odpowiedniego nadzoru i wsparcia ze strony człowieka system może generować nieprawdziwe informacje.
- **Mit 3: Generatywna AI działa bez ludzkiej interwencji.**
  AI może wykonywać zadania autonomicznie, chociaż wymaga nadzoru ze strony człowieka. Elementy takie jak cel biznesowy, kontekst oraz kluczowe przykłady z życia muszą być redagowane przez ludzi.
- **Mit 4: Generatywna AI zastąpi pracowników.**
  AI raczej wspiera pracowników, automatyzując rutynowe zadania, niż całkowicie ich zastępuje. Jest to narzędzie poprawiające wydajność, ale ludzie są potrzebni do tworzenia i kontroli wartości dodanej.
- **Mit 5: Implementacja AI jest prosta i szybka.**
  Wdrożenie generatywnej AI wymaga czasu i zasobów. Proces integracji AI z istniejącymi systemami, jej dostosowanie oraz zapewnienie zgodności z regulacjami i standardami etycznymi stanowi skomplikowane wyzwanie dla organizacji.

Zastosowanie AI w organizacji niesie zarówno wyzwania, jak i korzyści. Do głównych wyzwań należą konieczność posiadania specjalistycznej wiedzy, integracja technologii z istniejącymi procesami biznesowymi oraz zarządzanie zmianami w organizacji [Ledro i in., 2023]. Ponadto korzyści płynące z wykorzystania AI obejmują poprawę efektywności operacyjnej oraz tworzenie spersonalizowanych produktów i usług. W tym celu interdyscyplinarne podejście do projektów AI polega na integracji wiedzy oraz umiejętności z różnych dziedzin i specjalności. Dzięki współpracy ekspertów z różnych obszarów zainteresowania istnieje możliwość pełnego wykorzystania potencjału AI oraz osiągnięcia strategicznych celów firmy.

Kompleksowe podejście do nowych narzędzi ułatwia integrację AI w różnych aspektach działalności przedsiębiorstwa, co pozwala na efektywniejsze wykorzystanie danych, automatyzację procesów i optymalizację decyzji strategicznych. Dzięki temu organizacje mogą nie tylko zwiększać swoją konkurencyjność na rynku, ale także lepiej odpowiadać na zmieniające się potrzeby klientów i wyzwania technologiczne. Implementacja AI wymaga jednak odpowiedniej infrastruktury technologicznej, wsparcia specjalistów z różnych dziedzin oraz ciągłego szkolenia personelu, co jest niezbędne do pełnego wykorzystania potencjału oferowanego przez te nowoczesne technologie. Dzięki temu organizacje mogą znacznie lepiej zarządzać zmianami, angażując w proces większą liczbę interesariuszy oraz zwiększając skuteczność wykorzystania nowych rozwiązań [Ćwiertniak, 2024, s. 86].

W dojrzałych przedsiębiorstwach zastosowanie AI przebiega znacznie szybciej, ponieważ pracownicy są otwarci na zmiany, a zarządzanie innowacjami jest postrzegane jako czynnik długoterminowego rozwoju. Kultura innowacji nie tylko ułatwia wykorzystanie nowej technologii, ale także sprzyja innowacyjności oraz adaptacji nowych rozwiązań IT. W trakcie oceny zdolności organizacji do wdrażania AI istotne jest zatem zbadanie wybranych elementów, takich jak integracja AI z procesami biznesowymi oraz kultura organizacyjna. Te aspekty zostały dodatkowo omówione we wprowadzeniu do matrycy pomiaru (tabela 4.2).

4.4. Matryca pomiaru zdolności organizacji do wdrażania AI

Głównym wyzwaniem w dyfuzji projektów opartych na AI jest nie tylko adaptacja innowacji biznesowych do wymogów organizacji, ale także uzyskanie odpowiedniego poziomu akceptacji pracowników. Zintegrowana wiedza łącząca różne dziedziny, takie jak IT, HR i zarządzanie zmianą, może stymulować pracę, ułatwiając adaptację

nowych technologii. Taka współpraca pozwala uwzględnić zarówno techniczne, jak i społeczne aspekty wdrożenia AI.

Czynniki akceptacji innowacji opartej na AI związane są ze strategią, organizacją oraz procesem. Obejmują m.in. określenie nowych ról kierowniczych, tworzenie planów działania, projektowanie procesów, opanowanie narzędzi oraz reagowanie na zakłócenia w sytuacjach kryzysowych. Determinanty te można mierzyć w wybranych obszarach (tabela 4.1), takich jak: integracja narzędzi AI z długoterminowymi celami organizacji; dostosowanie struktury organizacyjnej do nowych wymagań wynikających ze sztucznej inteligencji; skuteczność dyfuzji innowacji oraz gotowość do reagowania na niespodziewane problemy związane z zastosowaniem nowej technologii.

Kluczowe jest więc zaangażowanie wszystkich poziomów zarządzania, aby zapewnić spójność i zgodność odnowionych procesów z misją oraz wizją organizacji. Dodatkowo regularne szkolenia oraz rozwój kompetencji pracowników są niezbędne do pełnego wykorzystania możliwości oferowanych przez sztuczną inteligencję.

Tabela 4.1. Wewnętrzne determinanty wdrażania AI w organizacji

Determinanty oceny	Komponenty oceny
Strategia innowacji	
Adaptacja strategii	znajomość dokumentów strategicznych, wskaźników innowacyjności
Dojrzałość projektowa	zdolność organizacji do identyfikowania czynników sukcesu
Organizacja i kultura innowacji	
Przywództwo	decyzyjność, kompetencje, umiejętność reagowania w sytuacjach kryzysowych
Współpraca	umiejętność pracy w grupie oraz wspólnego rozwiązywania problemów
Kultura	wyrozumiałość, tolerancja, zezwolenie na popełnianie błędów
Organizacja pracy	planowanie, koordynacja, udział personelu
Proces adaptacji	
Budżet	znajomość planu finansowego oraz inwestycyjnego
Metody wdrożeniowe	umiejętność wykorzystania istniejących metod organizacji i zarządzania
Kryteria oceny	jasno zdefiniowane, użyteczne, szeroko rozumiane determinanty oceny

Źródło: opracowanie własne na podstawie: Ćwiertniak [2024].

Zastosowanie sztucznej inteligencji (AI) w organizacji to złożony proces, który wymaga nie tylko odpowiednich metod działania, ale także przemyślanego zarządzania innowacjami. Sztuczna inteligencja ma potencjał do zmiany funkcjonowania organizacji, automatyzując procesy oraz wspierając podejmowanie decyzji. Jednak

skuteczna implementacja nowego systemu zależy od kilku innych czynników, takich jak: gotowość organizacji do zmian, zdolność systemu do integracji oraz kompetencje pracowników. W tym celu przedstawiono autorską matrycę oceny (tabela 4.2), która obejmuje trzy główne obszary pomiaru: funkcjonalność, wykonalność i opłacalność innowacji opartej na AI. Funkcjonalność odnosi się do jakości i innowacyjności rozwiązania, wykonalność obejmuje zdolność organizacji do skutecznego wdrożenia technologii, biorąc pod uwagę przygotowanie personelu, kulturę organizacyjną oraz wybrane narzędzia IT. Opłacalność dotyczy aspektów finansowych i ekonomicznych, takich jak rentowność projektu i jego wpływ na efektywność operacyjną.

Matryca pomiaru zdolności umożliwia organizacjom identyfikację mocnych i słabych stron w procesie wdrażania sztucznej inteligencji (AI). Wskazuje obszary, które wymagają optymalizacji, co jest niezbędne do maksymalizacji korzyści. Bez właściwej oceny kluczowych elementów projektu opartego na AI proces implementacji może napotkać liczne trudności. Prezentowany schemat stanowi zatem narzędzie wsparcia na każdym etapie zarządzania innowacjami – od ideacji, przez innowację, aż po dyfuzję. Ponadto przyjęta struktura pomiaru jest elastyczna i może być dostosowywana do indywidualnych potrzeb badanej organizacji.

Tabela 4.2. Ocena zdolności organizacji do wdrażania AI (OZO) – wybrane obszary

Kategoria zdolności organizacji do wdrażania AI		Ocena systemu	
		Pkt	Opis oceny
Ocena funkcjonalności innowacji AI			
Jakości innowacji – ZPI	znajomość metod planowania i kontroli jakości	1	brak działań związanych z oceną jakości
		2	realizowanie przez organizację w ograniczonym zakresie własnych badań jakości
		3	dział jakości, zajmujący się adaptacją rozwiązań spoza przedsiębiorstwa
Innowacyjność – ZIP	łatwość sprawnej, skutecznej, odpowiadającej oczekiwaniom jakościowym realizacji innowacji	1	brak działań dotyczących B+R
		2	prowadzenie przez organizację w ograniczonym zakresie własnych badań w celu stałej modernizacji i tworzenia nowych produktów
		3	prowadzenie przez organizację własnych badań i współpraca ze wyspecjalizowanymi jednostkami sektora B+R
System komunikacji – ZSI	opanowanie narzędzi IT, wykorzystanie baz danych oraz doświadczenia	1	gromadzenie i przechowywanie informacji na nośnikach papierowych
		2	istnienie w firmie wewnętrznej sieci informatycznej, obejmującej 30–50% pracowników
		3	istnienie w sieci wewnętrznej różnych baz danych oraz stosowane nowoczesnych systemów informatycznych

cd. tabeli 4.2

Kategoria zdolności organizacji do wdrażania AI		Ocena systemu	
		Pkt	Opis oceny
Ocena wykonalności innowacji AI			
Styl i system zarządzania – ZSZ	umiejętność wykorzystania istniejących standardów oraz metod organizacji i zarządzania	1	brak sprzedaży produktów innowacyjnych
		2	udział w sprzedaży produktów innowacyjnych wynoszący 5–49%
		3	udział w sprzedaży produktów innowacyjnych wynoszący więcej niż 50%
Kultura organizacyjna – ZKO	wyrozumiałość, tolerancja, możliwość popełniania błędów	1	brak wyraźnych przejawów kultury organizacyjnej oraz widocznych rezultatów działalności innowacyjnej
		2	widoczne artefakty zewnętrzne oraz umiejętność organizowania i realizacji prac zespołowych
		3	widoczne artefakty zewnętrzne i językowe, małe fluktuacje pracowników, dobra pozycja konkurencyjna przedsiębiorstwa
Ocena opłacalności innowacji AI			
Projekt inwestycyjny – ZEF	znajomość planu finansowego oraz inwestycyjnego	1	uzyskanie przez organizację znikomego dodatniego wyniku finansowego i nieplanowanie znaczących działań proinnowacyjnych
		2	uzyskanie przez organizację zysku wynoszącego 5–19% przychodów i przeznaczenie na działalność innowacyjną do 39% tego zysku
		3	uzyskanie przez organizację zysku wynoszącego ponad 20% przychodów i przeznaczanie na działalność innowacyjną ponad 40% tego zysku
Technologia – ZET	znajomość i umiejętność wykorzystania nowych technologii	1	brak służb i działań dotyczących oceny efektywności technicznej projektu,
		2	prowadzenie przez organizację w ograniczonym zakresie własnych badań efektywności technicznej projektu
		3	istnienie działu zajmującego się adaptacją rozwiązań przedsiębiorstwa do potrzeb wytwórczych organizacji

Źródło: opracowanie własne na podstawie: Ćwiertniak, Ćwiertniak [2022, s. 142–166].

Ogólny wskaźnik zdolności organizacji do wdrożenia AI w procesach biznesowych, przy uwzględnieniu współczynników wagowych, wyrażony według skali 1–3, można obliczyć według poniższego wzoru:

$$OZO = \frac{3(ZPI + ZIP + ZSI) + 2(ZEF + ZET) + ZSZ + ZKO}{15},$$

gdzie: *ZPI*, *ZIP*,..., *ZKO* – oznaczają konkretne wartości liczbowe na podstawie wcześniej wykonanej oceny poszczególnych kategorii (tabela 4.3).

Tabela 4.3. Wartości analizy funkcjonalności organizacji według kategorii

Kategoria	Wartość oceny	Charakterystyka
A	2,51–3,00	wielkość wzorcowa
B	2,01–2,50	stan wysokiej przydatności, wartość
C	1,51–2,00	stan użyteczny
D	1,00–1,50	nieużyteczność

Źródło: opracowanie własne na podstawie: Ćwiertniak, Ćwiertniak [2022, s. 142–166].

Wykorzystanie narzędzi sztucznej inteligencji stanowi złożony proces, który wpływa na różnorodne obszary całej organizacji [De Smet i in., 2024]. Wymaga on ścisłej współpracy między zespołem projektowym (kreatywnym i technologicznym) a zespołami odpowiedzialnymi za wdrożenie (produkcja, sprzedaż, marketing itp.). Ocena innowacji (zmiany) obejmuje w tym znaczeniu zarówno wymiar czasowy, jak i przestrzenny, co zakłada możliwość kontroli, która nie zawsze może być jednoznaczna. Takie działanie może prowadzić do napięć oraz niezadowolenia wśród pracowników oraz stać się źródłem potencjalnych konfliktów.

Podsumowanie

Na zakończenie należy stwierdzić, że wdrożenie sztucznej inteligencji w organizacjach ma potencjał, by przyczynić się do ich rozwoju. Jednakże skuteczność tego procesu zależy od przygotowania technologicznego, organizacyjnego oraz kulturowego. Nowe technologie, mimo że budzą obawy (m.in. związane z utratą miejsc pracy), jednocześnie tworzą nowe role dla pracowników, wspierające innowacje. Kluczowym czynnikiem sukcesu jest zatem zachowanie równowagi między technologią a zdolnością organizacji do adaptacji i zmiany.

Przedstawiona propozycja matrycy pomiaru zdolności organizacji do wdrażania sztucznej inteligencji umożliwia identyfikowanie kluczowych obszarów wymagających optymalizacji, minimalizowanie ryzyka oraz maksymalizowanie korzyści wynikających z innowacji. Taka ocena odgrywa kluczowe znaczenie dla podejmowania świadomych decyzji i wspierania innowacyjności. Organizacje powinny zatem doskonalić swoje systemy zarządzania innowacjami, w tym zintegrowane systemy wczesnego ostrzegania, aby efektywnie reagować na potencjalne zagrożenia.

Zastosowanie sztucznej inteligencji w firmie wymaga bowiem nie tylko odpowiedniej infrastruktury technologicznej, ale również inwestycji w kompetencje personelu

oraz tworzenia kultury sprzyjającej eksperymentom z nowymi systemami IT. Kluczowe aspekty, takie jak zarządzanie zmianami i akceptacja nowych rozwiązań przez pracowników, są ściśle związane z dojrzałością organizacyjną. Dopiero połączenie tych elementów z modelem zarządzania innowacjami pozwala na to, aby AI przyniosła trwałe korzyści i pozytywnie wpłynęła na konkurencyjność organizacji.

Bibliografia

Boston Consulting Group – BCG (2024). *How People Create and Destroy Value with Generative AI*, https://www.bcg.com/publications/2024/how-people-create-and-destroy-value-with-generative-ai (dostęp: 2.07.2024).

Carayannis, E.G., Samara, E.T. Bakouros, Y.L. (2015). *Innovation and Entrepreneurship: Theory, Policy and Practice*, https://dl.ojocv.gov.et/admin_/book/Innovation%20and%20Entrepreneurship_%20Theory,%20Policy%20and%20Practice%20 (%20PDFDrive%20).pdf (dostęp: 8.09.2024).

Cyberhaven (2024). *Trace Your Data to Protect It Like Never Before*, https://www.cyberhaven.com/ (dostęp: 2.07.2024).

Ćwiertniak, R. (2024). *Sztuczna inteligencja w organizacji. Innowacje biznesowe w praktyce*. Gliwice: Helion.

Ćwiertniak, R., Ćwiertniak, B. (2022). Ocena projektów innowacyjnych we wdrażaniu Smart City 3.0. W: *Inteligentne i zrównoważone miasta w teorii i praktyce zarządzania* (s. 142–166), G. Pawelska-Skrzypek, W. Blecharczyk (red.). Kraków: Wydawnictwo OGSMiE PAN.

De Smet, A., Durth, S., Hancock, B., Mugayar-Baldocchi, M., Reich, A. (2024). *The Human Side of Generative AI: Creating a Path to Productivity*. McKinsey & Company.

Desai, B., Patil, K., Patil, A., Mehta, I. (2023). Large Language Models: A Comprehensive Exploration of Modern AI's Potential and Pitfalls, *Journal of Innovative Technology*, 6(1), https://acadexpinnara.com/index.php/JIT/article/view/150 (dostęp: 10.03.2025).

Duarte, F. (2024). *Number of ChatGPT Users. Exploding Topics*, https://explodingtopics.com/blog/chatgpt-users (dostęp: 2.07.2024).

Grashof, N., Kopka, A. (2023). Artificial Intelligence and Radical Innovation: An Opportunity for All Companies?, *Small Business Economics*, 61, s. 771–797, https://link.springer.com/article/10.1007/s11187-022-00698-3 (dostęp: 8.09.2024).

Jiang, Y., Chen, C.C. (2018). Integrating Knowledge Activities for Team Innovation: Effects of transformational Leadership, *Journal of Management*, 44(5), s. 1819–1847, https://drive.google.com/file/d/1HGRPfNfSn2VUnLNcp1U-Hz7JyqsFktww/view (dostęp: 8.09.2024).

Kellogg, K.C., Sendak, M., Balu, S. (2023). Sztuczna inteligencja na pierwszej linii, *MIT Sloan Management Review Polska*, 15, s. 40–48.

Kelly, S., Kaye, S.A., Oviedo-Trespalacios, O. (2023). What Factors Contribute to the Acceptance of Artificial Intelligence? A Systematic Review, *Telematics and Informatics*, 77, https://www.sciencedirect.com/science/article/pii/S0736585322001587/pdf?trk=public_post_comment-text (dostęp: 8.09.2024).

Kozioł, L. (2013). Diagnoza innowacyjności przedsiębiorstw regionu Małopolski, *Zarządzanie i Finanse*, 4(1), s. 69–70.

Ledro, C., Nosella, A., Dalla Pozza, I. (2023). Integration of AI in CRM: Challenges and Guidelines, *Journal of Open Innovation: Technology, Market, and Complexity*, 9(4), https://www.sciencedirect.com/science/article/pii/S2199853123002536 (dostęp: 8.09.2024).

McKinsey & Company (2023). *People and Organizational Performance Practice: Human-Centered AI: The Power of Putting People First*, https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/human-centered-ai-the-power-of-putting-people-first (dostęp: 2.07.2024).

Miric, M. (2023). *Strategy, Not Technology, Is the Key to Winning with GenAI*, https://hbr.org/ 2023/12/strategy-not-technology-is-the-key-to-winning-with-genai?utm_medium=paidsearch&utm_source=google&utm_campaign=intlcontent_strategy&utm_term=Non- Brand&t-pcc=intlcontent_strategy&gad_source=1&gclid=CjwKCAjwm_SzBhAsEiwAXE2Cv4KUf4-6sHRW_uW-xn2DzE76YH8rsCqTRQD3TWiSjcTzQPQ_KEJIdxoCmpkQAvD_BwE (dostęp: 2.07.2024).

Nagarajan, T. (2023). *Genpact CEO Tiger Tyagarajan: AI Is Getting Good, But Still Can't Replace Human Curiosity*, https://hbr.org/2023/01/genpact-ceo-tiger-tyagarajan-ai-is-getting-good-but-still-cant-replace-human-curiosity (dostęp: 10.03.2025).

Porter, M.E., Heppelmann, J.E. (2015). How Smart, Connected Products Are Transforming Companies, *Harvard Business Review*, 93(10), s. 96–114, http://www.knowledgesol.com/uploads/2/4/3/9/24393270/hbr-how-smart-connected-products-are-transforming-companies.pdf (dostęp: 8.09.2024).

PwC (2018). *The $15 Trillion Question: Can You Trust Your AI?*, https://www.pwc.co.uk/services/risk/insights/explainable-ai.html (dostęp: 10.03.2025).

Sinha, V., Coffman, J., Fleming, R., Groves, B., Tejada, M.T. (2024). *Adapting Your Organization for Responsible AI*. Bain & Company.

Snyder, S.A., Velastegui, S. (2024). *Five Myths About Generative AI That Leaders Should Know*, https://knowledge.wharton.upenn.edu/article/five-myths-about-generative-ai-that-leaders-should-know/ (dostęp: 10.03.2025).

Waber, B., Fast, N.J. (2024). *Is GenAI's Impact on Productivity Overblown?*, https://hbr.org/2024/01/is-genais-impact-on-productivity-overblown (dostęp: 10.03.2025).

5.0

WYKORZYSTANIE SZTUCZNEJ INTELIGENCJI PRZEZ MAŁE I ŚREDNIE FIRMY W EUROPIE

Wojciech Nowakowski
Uniwersytet Warszawski

Streszczenie

Rozdział przedstawia badanie, którego celem było ustalenie poziomu wykorzystania sztucznej inteligencji (AI) przez małe i średnie przedsiębiorstwa (MŚP) w Europie. Z wykorzystaniem narzędzia analizy skupień i wykresu osuwiska wydzielono trzy skupienia, w których znajdują się państwa o zbliżonym do siebie poziomie wykorzystania rozwiązań AI przez MŚP. Wszystkie europejskie kraje zostały przypisane do jednego z trzech klastrów. Z analizy skupień wynika, że w klastrze liderów AI średnie wykorzystanie przez MŚP narzędzi AI spadło. Jest to jedyny klaster, w którym średnia z 2023 r. jest niższa niż z 2021. W pozostałych dwóch klastrach wzrosła średnia wykorzystania narzędzi AI przez małe i średnie firmy. Wzrost średnich w klastrze drugim jest większy niż w klastrze pierwszym. Z analizy wynika, że w państwach, które mają najwięcej do nadrobienia, wzrost jest szybszy. Skłania to do wniosku, że nie zawsze firmom opłaca się korzystać z nowych, rewolucyjnych rozwiązań.

Słowa kluczowe: sztuczna inteligencja, konkurencyjność, sektor małych i średnich przedsiębiorstw

Wprowadzenie

Rozwój sztucznej inteligencji (AI) stał się jednym z głównych tematów badań i dyskusji środowisk naukowych oraz biznesowych. Sztucznej inteligencji poświęca się wiele uwagi, opisuje się jej potencjalne zastosowanie, konsekwencje, wpływ na społeczeństwo czy też związane z nią zagrożenia i bariery. W dyskusji akademickiej zwrócono już uwagę, że AI może oddziaływać na wiele dziedzin i aspektów związanych z funkcjonowaniem gospodarki, pracą i życiem codziennym.

Wyniki badań pokazują również, że sztuczna inteligencja jest w stanie wpływać na rynek pracy poprzez zautomatyzowanie niektórych czynności, optymalizację zadań i poprawę efektywności pracy. Badacze wskazują również, że postęp w dziedzinie AI może skutkować istotnymi przemianami w zakresie podziału pracy i redystrybucji dochodów społecznych [Agrawal, Gans, Goldfarb, 2019; Arntz, Gregory, Zierahn, 2016; Atack, Margo, Rhode, 2019; Korinek, Stiglitz, 2017].

Zastosowanie przez firmy sztucznej inteligencji pozytywnie wpływa na poprawę konkurencyjności tych firm poprzez większe zdolności do realizowania określonych celów przedsiębiorstwa, jakimi są udział w rynku oraz zwiększenie obrotu i zysku [Hamilton, Tee, Maxwell, 2023; Kasior, 2024; Senadjki i in., 2023]. Poprawa konkurencyjności jest istotnym elementem rozwoju wszystkich przedsiębiorstw, dlatego też państwa na całym świecie stosują różne zachęty, aby firmy wdrażały i implementowały narzędzia AI. Sztuczna inteligencja to jeden z elementów sektora cyfrowego. Sektor ICT wraz z kluczowymi przejawami cyfrowej aktywności gospodarczej, np. takimi jak platformy internetowe, tworzy sektor cyfrowy, który rozwinął się niezwykle prężnie w ciągu ostatnich dwóch dekad. W wyniku rozszerzania się i nasilania pojedynczych procesów cyfryzacji we wszystkich obszarach życia gospodarczego i społecznego cała gospodarka ulega ucyfrowieniu [Śledziewska, Włoch, 2020]. Unia Europejska przeznacza znaczne środki finansowe na rozwój sektora cyfrowego. W budżecie na lata 2021–2027 przewidziano ponad 150 mld euro na rozwój jednolitego rynku, innowacji i gospodarki cyfrowej [European Council, 2024].

Celem niniejszej pracy jest ustalenie poziomu wykorzystywania przez małe i średnie firmy rozwiązań sztucznej inteligencji w różnych państwach europejskich z wykorzystaniem analizy klastrów oraz identyfikacja grup państw z podobnym wykorzystaniem narzędzi AI przez MŚP. Dodatkowo zbadane zostały odległości państw od centroidów poszczególnych klastrów. Ta analiza jest punktem wyjścia do dalszych badań, które mają na celu ukazanie głównych czynników warunkujących rozwój i integrację AI w europejskim sektorze małych i średnich firm.

Monitorowanie wykorzystania przez sektor MŚP narzędzi AI pomoże zobrazować i zrozumieć aktualne trendy w wykorzystaniu tej technologii do budowania konkurencyjności. Zrównoważony rozwój małych i średnich przedsiębiorstw jest kluczowy dla stałego rozwoju gospodarczego państw Unii Europejskiej. Wszystkie firmy w Unii Europejskiej powinny mieć możliwość uczciwej konkurencji i równego dostępu do narzędzi AI, tak aby nie powstały nierówności technologiczne, które mogłyby przełożyć się na nierówności ekonomiczne i społeczne. Jest to szczególnie ważne w kontekście małych i średnich przedsiębiorstw, odgrywających ważną rolę w rozwoju całej gospodarki w Unii Europejskiej [Lukács, 2005; Savlovschi, Robu, 2011].

5.1. Dane i metoda analizy skupień

Do przeprowadzenia badania wybrane zostały dane Eurostatu, które dostarczają informacji na temat procentowego udziału małych i średnich przedsiębiorstw, wykorzystujących co najmniej jedno narzędzie oparte na rozwiązaniach sztucznej inteligencji. W badaniu zastosowano definicję sztucznej inteligencji, którą proponuje Eurostat [2024a]: „Sztuczna inteligencja odnosi się do systemów, które wykorzystują technologie takie jak: eksploracja tekstu, wizja komputerowa, rozpoznawanie mowy, generowanie języka naturalnego, uczenie maszynowe, głębokie uczenie się w celu gromadzenia i/lub wykorzystywania danych do przewidywania, rekomendowania lub decydowania, z różnym poziomem autonomii, o najlepszym działaniu w celu osiągnięcia określonych celów".

Dane zostały umieszczone na stronie Eurostatu w zakładce „Nauka, technologia, cyfryzacja" w sekcji „Gospodarka i społeczeństwo cyfrowe" [Eurostat, 2024b]. Ich dostępność ograniczona była wyłącznie do dwóch lat pomiarowych: 2021 i 2023, co powoduje pewne trudności badawcze. Niepełność danych i niewielka liczba lat pomiarowych wynikają z nowości zagadnienia, co wskazuje, że narzędzia AI są stosowane przez małe i średnie firmy od niedawna. Dane dotyczyły 30 państw: wszystkich członków Unii Europejskiej oraz dodatkowo Turcji, Serbii i Norwegii. Taka dostępność danych pozwoliła na przeprowadzenie analizy skupień oraz porównanie powstałych klastrów.

Zebrane przez Eurostat dane zawierają pewne ograniczenia badawcze, takie jak możliwość wyznaczenia długookresowego trendu. Dodatkowo wskazują na potrzebę dalszego zgłębiania tematu, aby w większym stopniu wyjaśnić wykorzystanie rozwiązań AI przez małe i średnie firmy w Europie. W badaniu uwzględniono 30 państw, dla których wyznaczono odsetki przedsiębiorstw małych i średnich firm korzystających z AI w latach 2021 i 2023.

Tabela 5.1. Charakterystyka próby

Zmienna	Parametr	2021	2023	Różnica	Test	*p*-value
Odsetek MŚP korzystających ze sztucznej inteligencji	*N*	30	30	30	Wilcoxon	0,0093
	średnia	7,42	7,95	0,527		
	mediana	6,96	7,45	0,75		
	zakres	0,9–23,9	1,5–15,2	–8,7–3,5		

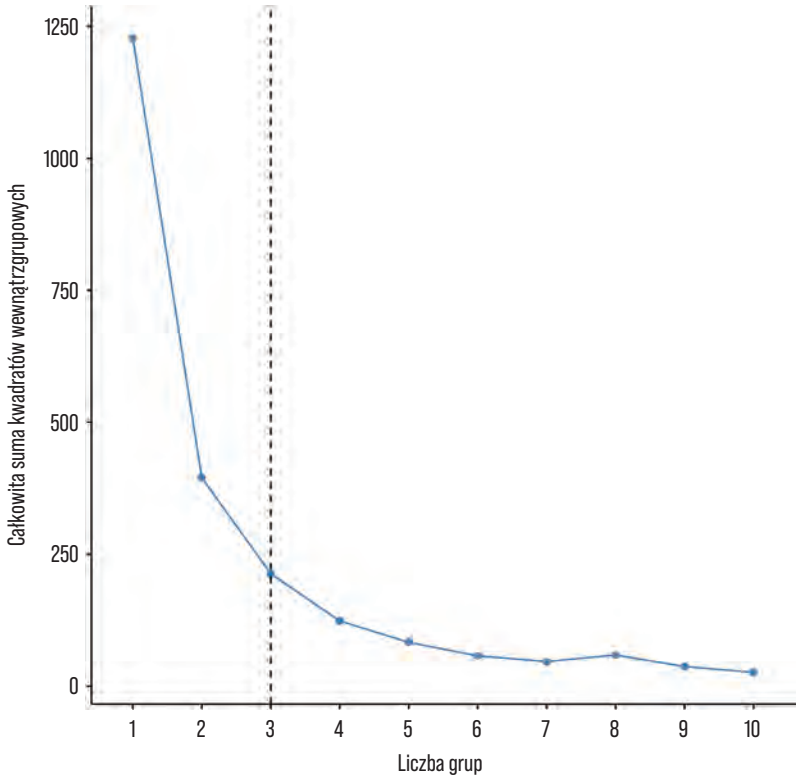
Źródło: opracowanie własne.

Zmienne, które zostały uwzględnione w badaniu, są pomiarami powtarzalnymi dla tych samych państw w latach 2021 i 2023 – aby zweryfikować istotność statystyczną zmiany poziomu wykorzystania AI przez MŚP wykonano nieparametryczny test Wilcoxona, ponieważ dane nie pochodziły z rozkładów normalnych. Przy poziomie istotności 0,05 wszystkie zmiany w poziomie wykorzystania sztucznej inteligencji przez MŚP są istotne statystycznie.

Do zmapowania poziomu wykorzystania sztucznej inteligencji w państwach europejskich użyto analizy skupień, wykonanej z wykorzystaniem metody k-średnich z algorytmem MacQueena [1967]. Analiza skupień to jedno z narzędzi analizy danych, które pozwalają na podzielenie danych na K niepustych, rozłącznych i podobnych pod względem wybranych zmiennych klastrów nazywanych skupieniami. Głównym celem tej analizy jest wykrycie z zbiorze danych tzw. naturalnych skupień, czyli skupień, które dają się w sensowny sposób interpretować [Wołyński, Górecki, 2013]. Istotą i zaletą analizy skupień jest zredukowanie liczby informacji do kilku podstawowych skupień, co pozwala na łatwe zorientowanie się w danym zjawisku oraz wyciągnięcie wniosków uogólniających [Pietrzykowski, Kobus, 2006]. Przy wykorzystaniu tej metody oblicza się środki klastrów, zwane centroidami, które są zróżnicowane, a obiekty wewnątrz grupy jednorodne. Aby wybrać optymalną liczbę klastrów, zastosowano metodę wykresu osypiska (*elbow method*), dzięki której udało się podzielić państwa na trzy klastry. Metoda ta pomaga określić optymalną liczbę klastrów na postawie wykresu zależności między liczbą klastrów a sumą kwadratów odchyleń wewnątrz klastrów. Punkt na wykresie, gdzie zauważalna jest znaczna zmiana nachylenia wykresu (*elbow*), wskazuje, że dodanie następnych klastrów nie przyniesie już istotnej poprawy jakości podziału. W badaniu zdecydowano się na wybór trzech klastrów, ponieważ taka ich liczba umożliwiła wyodrębnienie możliwie jednorodnych klastrów, charakteryzujących się centroidami (wartościami średnimi, środkowymi zmiennych, na których oparto analizę skupień). W przypadku wyodrębnienia czterech klastrów dodatkowy klaster składałby się jedynie z Danii, która wykazuje odstający poziom w analizie.

Oznaczałoby to pominięcie Danii w analizie, a ponieważ jest to kraj należący do UE, nie chciano go wykluczać. Rysunek 5.1 przedstawia osypiska bazujące na całkowitej sumie kwadratów wewnątrzgrupowych.

Rysunek 5.1. Wykres osypiska



Źródło: opracowanie własne.

Trzeba zwrócić uwagę na fakt, że wykorzystanie sztucznej inteligencji to stosunkowo nowe rozwiązania, dlatego też przedsiębiorcy uczą się jeszcze stosować je do różnych procesów. Występują też zauważane różnice pomiędzy zastosowaniem sztucznej inteligencji w zależności od branży. Rozwój wykorzystania sztucznej inteligencji prowadzi do adaptacji technologii w klasycznych sektorach gospodarki w całym łańcuchu wartości i przekształca je, prowadząc do algorytmizacji niemal wszystkich funkcjonalności, od logistyki po zarządzanie firmą [Berdiyorova, Akhtamova, Ganiev, 2021]. Możliwe, że wykorzystanie sztucznej inteligencji zależy od wielkości gospodarki, zróżnicowania sektorów gospodarki, rozwoju poszczególnych branż oraz dostępnych zasobów finansowych, co wymagałoby dalszego zbadania.

5.2. Wyniki

Przeprowadzona analiza skupień pozwoliła podzielić wszystkie badane państwa na trzy klastry ze względu na wykorzystanie narzędzi AI przez MŚP. Poniżej zaprezentowano centroidy dla wszystkich klastrów państw.

Centroid klastrowy, będący punktem centralnym klastra, reprezentuje średni profil wykorzystania AI w danym klastrze. Tabela 5.2 przedstawia centroidy dla trzech klastrów w latach 2021 i 2023.

Tabela 5.2. Centroidy w poszczególnych klastrach

Klaster	2021	2023
1	9,44	10,31
2	3,53	4,46
3	16,45	14,53

Źródło: opracowanie własne.

Interpretacja przedstawionych wyników centroidów dla trzech klastrów w latach 2021 i 2023 wskazuje na zróżnicowaną dynamikę zmian w wykorzystaniu narzędzi sztucznej inteligencji przez małe i średnie przedsiębiorstwa w różnych klastrach państw. Wyniki są zaskakujące, ponieważ obecnie następuje bardzo szybki rozwój AI, dodatkowo dużo narzędzi sztucznej inteligencji jest bezpłatnych, a do tego zyskuje bardzo szybko nowych użytkowników, dlatego interesujący wydaje się spadek wartości centroidu klastra 3.

Klaster 3 odnotował spadek wartości centroidu z 16,45 w 2021 r. do 14,53 w 2023. To jedyny klaster, w którym zanotowano spadek. Dodatkowo należy zwrócić uwagę, że jest to klaster liderów AI, państw, w których największy odsetek małych i średnich przedsiębiorstw korzysta z narządzi AI.

W klastrze 1 centroid wzrósł z wartości 9,44 w 2021 r. do 10,31 w 2023 r., co wskazuje na wzrost zastosowania narzędzi AI przez MŚP.

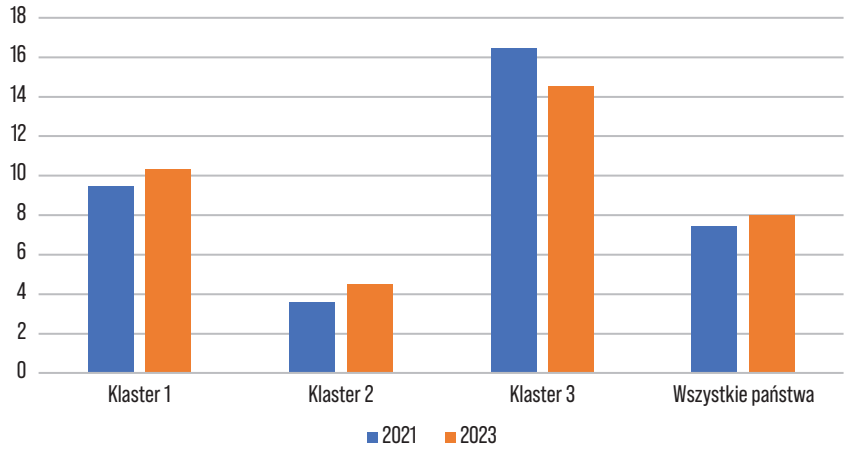
Klaster 2 także wykazał wzrost centroidów, z 3,53 w 2021 r. do 4,46 w 2023 r., co świadczy o zwiększonym wykorzystaniu AI, choć z niższego poziomu startowego w porównaniu z klastrem 1.

Poniżej zaprezentowano również średnie dla poszczególnych klastrów oraz zmiany, które zaszły w wynikach między 2021 a 2023 r.

Na rysunku 5.2 przedstawiono średnie wykorzystanie narzędzi AI przez małe i średnie firmy w dwóch obserwowanych latach. Średnie przedstawiono za pomocą

analizy skupień klastrów dla wszystkich wyróżnionych oraz dodatkowo dla wszystkich badanych państw łącznie.

Rysunek 5.2. Średnie wykorzystanie narzędzi AI w MŚP w poszczególnych klastrach (%)



Źródło: opracowanie własne.

Analiza danych z lat 2021 i 2023 wykazała, że średnie wykorzystanie narzędzi AI w klastrze 3 spadło z 16,45% do 14,525%, co reprezentuje spadek o 1,925 p.p. Warto zauważyć, że jest to jedyne skupienie, w którym odnotowano spadek. Klaster 1 wykazał wzrost ze średniej 9,44% do 10,3%, co oznacza wzrost o 0,86 p.p. Również klaster 2 odnotował wzrost wykorzystania AI, z 3,53% do 4,46%, co przekłada się na zwiększenie o 0,93 p.p. Przy rozpatrywaniu wszystkich krajów jako jednej kategorii średnie wykorzystanie AI przez małe i średnie przedsiębiorstwa wzrosło z 7,42% w 2021 r. do 7,95% w 2023, co wskazuje na umiarkowany wzrost o 0,527 p.p. Wśród 30 analizowanych państw tylko siedem zanotowało spadek wykorzystywanych rozwiązań sztucznej inteligencji w małych i średnich firmach. Największy spadek, aż o 8,7 p.p., dotyczy Danii, która pomimo tak dużego spadku dalej jest w Europie liderem w wykorzystaniu sztucznej inteligencji przez małe i średnie firmy. Drugi pod względem wielkości spadek, o 1,6 p.p., zanotowała Norwegia, zaś trzeci dotyczy Włoch, w których liczba małych i średnich firm wykorzystujących rozwiązania sztucznej inteligencji zmniejszyła się o 1,2 p.p. Przyczyny spadków w tych krajach mogą być zróżnicowane. Możliwe, że rynek osiągnął nasycenie w zakresie zastosowań sztucznej inteligencji (AI) przez sektor małych i średnich przedsiębiorstw (MŚP), co spowodowało stagnację w dalszym rozwoju. Inne czynniki mogą obejmować zamknięcie firm korzystających z AI z powodu trudności finansowych lub przesunięcie priorytetów w kierunku innych

innowacji, takich jak transformacja energetyczna. Analiza barier ograniczających wdrażanie AI przez MŚP jest istotnym zagadnieniem, które wymaga dalszych badań.

Wzrosty w wykorzystaniu AI, w klastrach 1 i 2, mogą być rezultatem kilku czynników. Przede wszystkim przedsiębiorstwa starają się dogonić liderów rynkowych, którzy już osiągnęli przewagę dzięki sztucznej inteligencji. Możliwe, że nowo powstałe firmy, chcąc osiągnąć przewagę konkurencyjną i wyróżnić się na tle konkurencji, od razu zaczynają od wdrażania AI. Wzrosty mogą także wynikać z rosnącej świadomości firm co do korzyści płynących z zastosowania sztucznej inteligencji.

Przyczyny wzrostu i spadków wykorzystywania AI przez sektor MŚP powinny być przedmiotem dalszych badań i dyskusji naukowych.

Zgromadzone dane dotyczące lat 2021 i 2023 pokazują spadek zaangażowania w technologie AI w klastrze krajów, które pierwotnie charakteryzowały się najwyższym odsetkiem implementacji tych rozwiązań. Jest to zjawisko wymagające głębszej analizy, mogące sugerować dojście do punktu nasycenia, problemy ze skalowalnością rozwiązań AI czy konieczność dokonywania efektywniejszych inwestycji w tej sferze. Zmiany w klastrze 1 i 2 pokazują, że rośnie odsetek małych i średnich firm wykorzystujących rozwiązania AI w obu tych klastrach

Na podstawie analizy skupień kraje zostały podzielone na trzy klastry, które różnią się poziomem wykorzystania rozwiązań sztucznej inteligencji przez małe i średnie firmy. Klaster 1 to średniacy AI, klaster 2 stanowią początkujący AI, klaster 3 to liderzy AI.

Klaster 1 – średniacy AI – to kraje o średnim poziomie wykorzystania sztucznej inteligencji w małych i średnich przedsiębiorstwach. Zalicza się do niego 11 państw, są to: Belgia, Niemcy, Irlandia, Hiszpania, Chorwacja, Malta, Austria, Portugalia, Słowenia, Szwecja, Norwegia.

Średniacy AI to klaster państw, w których około 10% małych i średnich przedsiębiorstw korzysta z chociaż jednego rozwiązania sztucznej inteligencji. Liderami tego klastra są: Belgia, Niemcy oraz Irlandia.

Klaster 2 – początkujący AI – to kraje o niskim poziomie wykorzystania sztucznej inteligencji w małych i średnich przedsiębiorstwach. Należy do niego 15 państw: Bułgaria, Czechy, Estonia, Grecja, Francja, Włochy, Cypr, Łotwa, Litwa, Węgry, Polska, Rumunia, Słowacja, Serbia, Turcja.

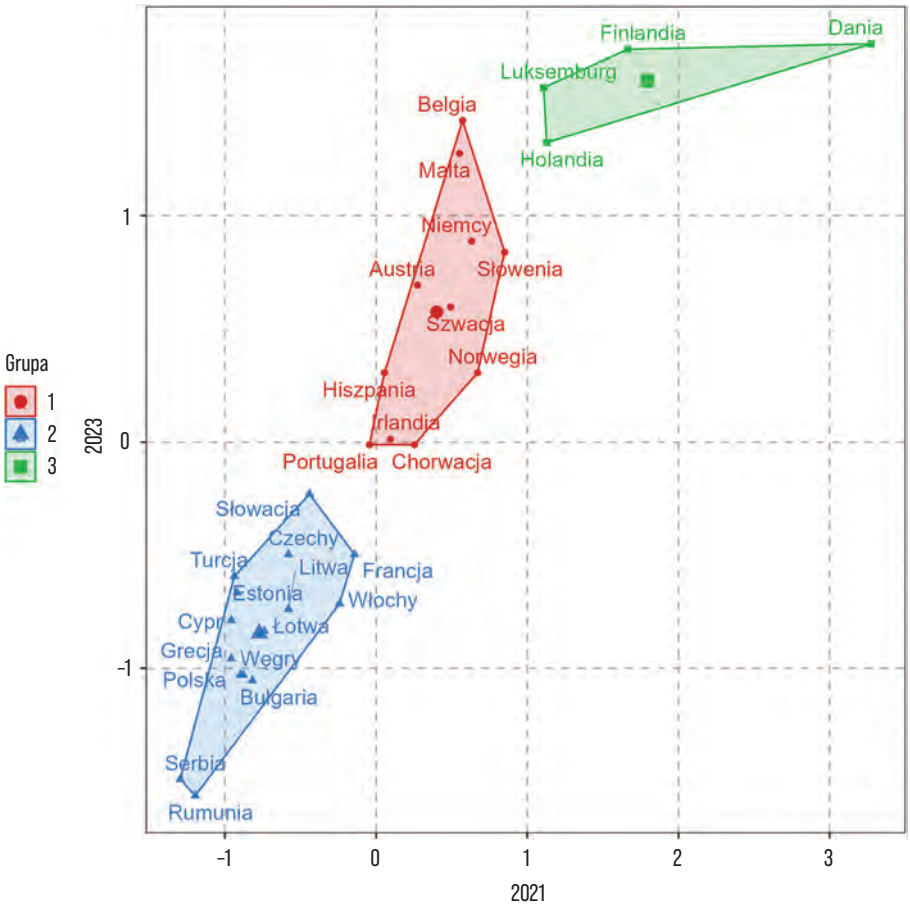
Początkujący AI to najliczniejszy klaster, w którym jedynie około 4% małych i średnich przedsiębiorstw korzysta z chociaż jednego rozwiązania sztucznej inteligencji.

Klaster 3 – liderzy AI – to kraje o wysokim poziomie wykorzystania sztucznej inteligencji w małych i średnich przedsiębiorstwach. Należą do nigo cztery państwa: Dania, Luksemburg, Holandia, Finlandia.

Liderzy AI to kraje, w których około 15% małych i średnich firm korzysta z rozwiązań sztucznej inteligencji. Liderem klastra jest Dania, gdzie w 2023 r. 15,2% małych i średnich firm korzystało z przynajmniej jednego rozwiązania sztucznej inteligencji.

Na rysunku 5.3 zaprezentowano rozmieszczenie państw w poszczególnych klastrach skupień. Wśród liderów AI wyraźnie wyróżniają się dane z 2021 r. dotyczące Danii, gdzie w omawianym czasie aż 23,9% małych i średnich firm korzystało z rozwiązań sztucznej inteligencji. Zauważyć można też ogromne dysproporcje w liczbie państw składających się na dany klaster. Na 30 przedstawionych państw jedynie cztery znajdują się w klastrze 3 – liderzy AI.

Rysunek 5.3. **Położenie państw w dwuwymiarowej przestrzeni wyznaczonej przez znormalizowane wartości pomiarów w latach 2021 i 2023**



Źródło: opracowanie własne.

W świetle bieżącej analizy dotyczącej implementacji rozwiązań sztucznej inteligencji przez małe i średnie przedsiębiorstwa w różnych państwach europejskich już na wstępnym etapie wprowadzania tych rozwiązań widoczne są wyraźne dysproporcje. Pomimo relatywnej nowości tematu i rosnącej powszechności sztucznej inteligencji badania wskazują na niejednorodne wyniki wykorzystania sztucznej inteligencji przez małe i średnie przedsiębiorstwa, co może odzwierciedlać zarówno zróżnicowane ścieżki rozwoju technologicznego, jak i nierównomierność dostępu do zasobów i wiedzy, niezbędnych do skutecznej integracji rozwiązań AI w MŚP.

Skład poszczególnych klastrów jest interesujący. W państwach zachodniej i północnej Europy z rozwiązań sztucznej inteligencji korzysta zdecydowanie więcej małych i średnich firm niż w państwach południowej i wschodniej Europy. W następnym badaniu należy poszukać determinant wpływających na przyjęcie rozwiązań sztucznej inteligencji w małych i średnich firmach w państwach europejskich.

W celu zbadania odległości poszczególnych państw od centroidów wykorzystano metodę euklidesową, która polega na obliczeniu odległości geometrycznej w przestrzeni wielowymiarowej:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2},$$

gdzie: x_1, y_1 to współrzędne pierwszego punktu, a x_2, y_2 to współrzędne drugiego punktu.

Wyniki odległości przedstawiono w wartościach bezwzględnych.

Odległość poszczególnych państw od centroidu świadczy o wielkości odsetka MŚP, które wykorzystują sztuczną inteligencję. Kraje położone najbliżej centroidu można uznać za typowych przedstawicieli danego klastra.

Klaster 1 to kraje o średnim poziomie wykorzystania narzędzi AI przez MŚP. W tym klastrze Szwecja znajduje się najbliżej centroidu, co wskazuje, że wykorzystanie AI przez szwedzkie MŚP jest najbardziej reprezentatywne dla tego klastra. Kraje takie jak Belgia i Portugalia są dalej od centroidu, co sugeruje, że w tych krajach wykorzystanie narzędzi AI przez MŚP jest mniej typowe niż w innych państwach w klastrze.

W klasterze 2, w którym znajdują się kraje z najniższym wykorzystaniem narzędzi AI wśród MŚP, Łotwa znajduje się najbliżej centroidu, stosowanie AI w tym kraju jest bliskie średniej dla klastra, podczas gdy Serbia jest najdalej, co może oznaczać, że serbskie MŚP są najmniej zaangażowane w integrację narzędzi AI.

5.3. Dyskusja

Analiza skupień pozwoliła na podzielenie państw europejskich ze względu na wykorzystanie narzędzi sztucznej inteligencji przez małe i średnie firmy i wydzielenie trzech klastrów (początkujący AI, średniacy AI, liderzy AI). Następnie analiza średnich i centroidów w klastrach pokazała zmiany, jakie zaszły pomiędzy 2021 a 2023 r. w każdym z wyodrębnionych klastrów. Niezbędne są dalsze badania, które wskażą, dlaczego następują zmiany w odsetku MŚP wykorzystujących narzędzia AI w Europie. Dodatkowo należy poświęcić uwagę kwestii determinant, wpływających na przyjmowanie przez sektor MŚP nowych narzędzi AI.

Zaprezentowane w rozdziale wyniki są zgodne z najnowszymi publikacjami i badaniami naukowymi. Z opublikowanego w 2020 r. badania wynika, że państwami europejskimi, które w największym stopniu wprowadziły AI w usługach publicznych, są Holandia – 20 rozwiązań AI oraz Dania – 19 rozwiązań AI. Badaniu poddano łącznie 15 państw, pozostałe państwa wdrożyły około 3 rozwiązań AI [Misuraca, Noordt, Boukli, 2020].

W innym badaniu, poświęconym dojrzałości cyfrowej europejskich przedsiębiorstw, podzielono państwa europejskie na cztery klasy. Dania, Malta i Finlandia znalazły się w klasie „ekspertów" pod względem dojrzałości biznesu cyfrowego. Klasa „doświadczonych" obejmowała Belgię, Niemcy, Irlandię, Hiszpanię, Chorwację, Włochy, Holandię, Słowenię i Szwecję, a klasa „średniozaawansowanych": Czechy, Estonię, Francję, Cypr, Litwę, Luksemburg, Austrię, Polskę, Portugalię i Słowację. Do klasy „początkujących" z kolei zaliczone zostały: Węgry, Rumunia, Bułgaria, Grecja i Łotwa [Tutak, Brodny, 2021].

W badaniu poświęconym transformacji cyfrowej, opartej na technologiach AI w organizacjach Unii Europejskiej, wskazuje się, że państwa, które w największym stopniu wykorzystują technologię AI, to: Finlandia, Szwecja, Dania oraz Holandia. Z kolei państwami, które w najmniejszym stopniu wykorzystują AI w organizacjach, są: Rumunia, Serbia, Bułgaria i Czarnogóra [Mihai, Aleca, Gheorghe, 2023].

Wyniki zaprezentowane w powyższym rozdziale w znacznym stopniu pokrywają się z wymienionymi badaniami i wskazują podobny poziom wykorzystania AI przez poszczególne państwa Unii Europejskiej.

Należy też zwrócić uwagę, że wszystkie badania oraz analiza zawarta w rozdziale w dużym stopniu pokrywają się z Indeksem Gospodarki Cyfrowej i Społeczeństwa (DESI), który jest corocznym raportem publikowanym przez Komisję Europejską. Różnice w rankingu są oczywiście związane z metodologią badań, tj. z różnymi wskaźnikami

przyjętymi do analizy oraz z faktem, że indeks DESI dotyczy całych społeczeństw krajów, a nie tylko przedsiębiorstw [Tutak, Brodny, 2021].

Analiza ta zwraca uwagę na złożoność wyzwań, związanych z wdrażaniem nowych technologii w sektorze MŚP, dla którego czynniki, takie jak infrastruktura techniczna, kapitał ludzki, wsparcie rządowe i regulacje, będą miały wpływ na wykorzystywanie narzędzi AI.

Rozwój AI staje się kluczowym elementem konkurencyjności na rynku globalnym, dlatego istotne jest zrozumienie czynników wpływających na przyjmowanie przez MSP rozwiązań AI. Długookresowa polityka państw UE powinna opierać się na tworzeniu równych szans do wykorzystywania rozwiązań AI przez MŚP we wszystkich krajach Unii Europejskiej. W dłuższej perspektywie wpłynie to na poprawę konkurencyjności firm europejskich, a w konsekwencji zwiększą one swój udział w rynkach i rentowności, co przyczyni się do rozwoju społecznego i gospodarczego Europy.

Zauważalny ogólny wzrost średniego wykorzystania AI we wszystkich krajach wskazuje na postępującą inklinację do wykorzystywania technologii AI w sektorze MŚP. Mimo to konieczne są dalsze badania, które potwierdzą lub odrzucą dotychczasowe obserwacje. Przyglądanie się wykorzystaniu rozwiązań AI przez MŚP w Europie pozwoli na wyznaczenie trendu. Dalsze badania pomogą zrozumieć i scharakteryzować procesy i determinanty przemawiające za technologiczną rewolucją, jaką jest wykorzystywanie AI w pracy. Analizy te mogą przyczynić się do identyfikacji barier technologicznych, ekonomicznych i organizacyjnych, które przedsiębiorcy napotykają aktualnie na swojej drodze. Informacje o dalszym rozwoju AI w MŚP pomogą w monitorowaniu stanu wykorzystania tych narzędzi przez firmy oraz będą wskazówką przy tworzeniu polityki wspierającej korzystanie z nich.

Podsumowanie

Przeprowadzona analiza pokazuje ilościowe zmiany, które zaszły w wykorzystaniu narzędzi AI przez europejskie MŚP. Podział na klastry pozwolił na zgrupowanie państw o zbliżonym poziomie wykorzystywania rozwiązań narzędzi AI przez małe i średnie przedsiębiorstwa.

Z analizy skupień wynika, że w klastrze liderów AI średnie wykorzystanie przez MŚP narzędzi AI spadło. To jedyny klaster, w którym średnia z 2023 r. jest niższa niż z 2021. Spadek ten wynika ze zmniejszenia się liczby MŚP, które korzystają z narzędzi AI, u dwóch liderów tego klastra: Danii i Finlandii. W pozostałych dwóch klastrach wzrosła średnia wykorzystania narzędzi AI przez małe i średnie firmy. Wzrost śred-

nich w klastrze początkujących AI jest większy niż w klastrze średniaków AI. Z analizy wynika, że w państwach, które mają najwięcej do nadrobienia, wzrost jest szybszy. Skłania to do wyciągnięcia wniosku, że może nie zawsze firmom opłaca się korzystać z nowych, rewolucyjnych rozwiązań. Być może w wykorzystaniu narzędzi AI lepiej przeprowadzać transformację powoli, gdyż z czasem niepewność i ryzyko związane z legislacją i kosztami będą już znane i wytyczone przez inne firmy.

Interesujące jest to, że kraje zachodniej i północnej Europy są aktywniejsze w wykorzystaniu AI niż kraje południowej i wschodniej Europy, co może wskazywać na geograficzne i gospodarcze czynniki wpływające na przyjęcie innowacji technologicznych.

W konkluzji dane te dostarczają wskazówek dotyczących stanu wykorzystania AI przez MŚP w różnych częściach Europy i mogą służyć jako punkt wyjścia do głębszej analizy determinant takiej dyfuzji technologicznej.

Badanie należy powtórzyć, gdy pojawią się kolejne dane, na podstawie których będzie można wyznaczyć trendy. Interesujące wydaje się również badanie ekonomicznych i społecznych czynników, które wpływają na chęć korzystania z narzędzi AI przez małe i średnie przedsiębiorstwa w poszczególnych państwach europejskich.

Bibliografia

Aghion, P., Jones, B.F., Jones, C.I. (2017). *Artificial Intelligence and Economic Growth*. Technical Report. NBER Working Paper 23928. Cambridge, MA: National Bureau of Economic Research.

Agrawal, A., Gans, J.S., Goldfarb, A. (2019). Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction, *Journal of Economic Perspectives*, 33, s. 31–50.

Acemoglu, D., Restrepo, P. (2019). Automation and New Tasks: How Technology Displaces and Reinstates Labor, *Journal of Economic Perspectives*, 33, s. 3–30.

Arntz, M., Gregory, T., Zierahn, U. (2016). *The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis*. Technical Report. OECD Social, Employment and Migration Working Papers 189. Paris: OECD Publishing.

Atack, J., Margo, R.A., Rhode, P.W. (2019). "Automation" of Manufacturing in the Late Nineteenth Century: The Hand and Machine Labor Study, *Journal of Economic Perspectives*, 33(2), s. 51–70.

Berdiyorova, I., Akhtamova, P., Ganiev, I.M. (2021). Artificial Intelligence in Various Industries, *Development Issues of Innovative Economy in the Agricultural Sector*, s. 750–757.

Brynjolfsson, E., Rock, D., Syverson, C. (2018). Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics. W: *Economics of Artificial Intelligence: An Agenda*, A. Agrawal, J. Gans, A. Goldfarb (Eds.). Chicago: University of Chicago Press.

Brynjolfsson, E., Rock, D., Syverson, C. (2019). Artificial Intelligence and the Modern Productivity Paradox, *The Economics of Artificial Intelligence: An Agenda*, 23, s. 23–57.

Buarque, B.S., Davies, R.B., Hynes, R.M., Kogler, D.F. (2020). Cambridge Journal of Regions, *Economy and Society*, 13(1), s. 175–192.

Cockburn, I.M., Henderson, R., Stern, S. (2018). *The Impact of Artificial Intelligence on Innovation*. Technical Report. NBER Working Paper 24449. Cambridge, MA: National Bureau of Economic Research.

European Council (2024). *The EU Long-Term Budget*, https://www.consilium.europa.eu/pl/policies/eu-long-term-budget (dostęp: 11.06.2024).

Eurostat (2024a). *Glossary: Artificial Intelligence (AI)*, https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Artificial_intelligence_(AI) (dostęp: 10.06.2024).

Eurostat (2024b). *Procentowy udział małych i średnich przedsiębiorstw wykorzystujących co najmniej jedno narzędzie oparte na rozwiązaniach sztucznej inteligencji*, https://ec.europa.eu/eurostat/databrowser/view/ISOC_EB_AI__custom_15770975/default/table?lang=en (dostęp: 9.04.2024)

Frey, C.B., Osborne, M.A. (2017). The Future of Employment: How Susceptible Are Jobs to Computerization?, *Technological Forecasting and Social Change*, 114, s. 254–280.

Goldfarb, A., Trefler, D. (2018). AI and International Trade. W: *The Economics of Artificial Intelligence: An Agenda*, A. Agrawal, J. Gans, A. Goldfarb (Eds.). Chicago: University of Chicago Press.

Grguric, A., Vlacic, E., Drvenkar, N. (2020). Assessing Firms' Competitiveness and Technological Advancement By Applying Artificial Intelligence as a Differentiation Strategy-A Proposed Conceptual Mode. W: *Economic and Social Development*: *Book of Proceedings* (s. 43–61), M, Milkovic, K, Hammes, O. Bakhtina (Eds.). Varazdin: Varazdin Development and Entrepreneurship Agency.

Hamilton, J.R., Tee, S.W., Maxwell, S.J. (2023). *AI and Firm Competitiveness*. ICEB 2023 Proceedings (Chiayi, Taiwan).

Korinek, A. Stiglitz, J.E. (2017). *Artificial Intelligence and Its Implications for Income Distribution and Unemployment*. Technical Report. NBER Working Paper 24174. Cambridge, MA: National Bureau of Economic Research.

Kasior, K. (2024). Sztuczna inteligencja w przedsiębiorstwach przemysłu spożywczego – potencjał i wyzwania, *Przemysł Spożywczy*, 78(2), s. 12–17.

Lukács, E. (2005). The Economic Role of SMEs in World Economy, Especially in Europe, *European Integration Studies*, 4(1), s. 3–12.

MacQueen, J. (1967). Some Methods for Classification and Analysis of Multivariate Observations. W: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, *1: Statistics* (vol. 5, s. 281–297). University of California (Press).

Mihai, F., Aleca, O.E., Gheorghe, M. (2023). Digital Transformation Based on AI Technologies in European Union Organizations, *Electronics*, 12(11), 2386.

Misuraca, G., van Noordt, C., Boukli, A. (2020). The Use of AI in Public Services: Results from a Preliminary Mapping Across the EU. W: *Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance* (s. 90–99). DOI: 10.1145/3428502.3428513.

Migdał-Najman, K., Najman, K. (2013). Analiza porównawcza wybranych metod analizy skupień w grupowaniu jednostek o złożonej strukturze grupowej, *Zarządzanie i Finanse*, 11(3, cz. 2), s. 179–194.

Muller, P., Caliandro, C., Peycheva, V., Gagliardi, D., Marzocchi, C., Ramlogan, R., Cox, D. (2015). Annual Report on European SMEs, *European Commission*, 5(1), s. 36–48.

Pietrzykowski, R., Kobus, P. (2006). Zastosowanie modyfikacji metody k-średnich w analizie portfelowej, *Zeszyty Naukowe SGGW. Ekonomika i Organizacja Gospodarki Żywnościowej*, 60, s. 301–308.

Savlovschi, L.I., Robu, N.R. (2011). The Role of SMEs in Modern Economy. *Economia. Seria Management*, 14(1), s. 277–281.

Senadjki, A., Ogbeibu, S., Mohd, S., Hui Nee, A.Y., Awal, I.M. (2023). Harnessing Artificial Intelligence for Business Competitiveness in Achieving Sustainable Development Goals, *Journal of Asia-Pacific Business*, 24(3), s. 149–169. DOI: 10.1080/10599231.2023.2220603.

Skare, M., de Obesso, M., Ribeiro-Navarrete, S. (2023). Digital Transformation and European Small and Medium Enterprises (SMEs): A Comparative Study Using Digital Economy and Society Index Data, *International Journal of Information Management*, 68, 102594.

Śledziewska, K., Włoch, R. (2020). *Gospodarka cyfrowa. Jak nowe technologie zmieniają świat?*. Warszawa: Wydawnictwo Uniwersytetu Warszawskiego.

Tutak, M., Brodny, J. (2022). Business Digital Maturity in Europe and Its Implication for Open Inovation, *Journal of Open Innovation: Technology, Market, and Complexity*, 8(1), s. 27.

Wołyński, W., Górecki, T. (2013). *Analiza skupień*. Poznań: Wydział Matematyki i Informatyki UAM.

II

TECHNOLOGIE, DANE I ANALITYKA SZTUCZNEJ INTELIGENCJI

6.0

WYZWANIA ZWIĄZANE Z DOSZKALANIEM DUŻYCH MODELI JĘZYKOWYCH – PERSPEKTYWA EKONOMETRII

Igor Kapiczyński
Szkoła Główna Handlowa w Warszawie

Dawid Zdanowicz
Szkoła Główna Handlowa w Warszawie

Daniel Kaszyński
Szkoła Główna Handlowa w Warszawie

Petro Vavryk
Taras Shevchenko National University of Kyiv

Streszczenie

Celem rozdziału jest przedstawienie konsekwencji, jakie mogą powstać w wyniku doszkalania dużych modeli językowych, oraz ich przykładowych miar ewaluacji. Fenomen, który zidentyfikowano w rozdziale na przykładzie rzeczywistych danych, znany jest w literaturze jako problem katastrofalnego zapominania (*catastrophic forgetting*). Opisane zjawisko prowadzi do spadku jakości modelu w obszarach niezależnych od obszaru dotrenowanego. Celem tekstu jest wskazanie na rozwój modeli ekonometrycznych, skutkujący obecnie powstaniem modeli klasy LLM oraz zwrócenie uwagi na konsekwencje doszkalania tych modeli, które wpływają na ich użyteczność.

Słowa kluczowe: sztuczna inteligencja, duże modele językowe, doszkalanie dużych modeli językowych, katastrofalne zapominanie

Wprowadzenie

Pierwsze modele ekonometryczne – współcześnie często zaliczane do modeli uczenia maszynowego – są znane od połowy XX w. dzięki pracom Ragnara Frischa, laureata Nagrody Nobla z 1969 r. W swoim artykule pt. *O problemie czystej ekonomii* wskazuje on na ekonometrię jako nową dyscyplinę naukową [Frisch, 1926, s. 1]: „Na pograniczu matematyki, statystyki i ekonomii znajduje się nowa dyscyplina, którą z braku lepszej nazwy można nazwać ekonometrią. Celem ekonometrii jest poddanie abstrakcyjnych praw teoretycznej ekonomii politycznej lub «czystej» ekonomii eksperymentalnej i numerycznej weryfikacji, a tym samym przekształcenie czystej ekonomii, w miarę możliwości, w naukę w ścisłym tego słowa znaczeniu".

Pierwotnie narzędziami ekonometrii były modele parametryczne, takie jak: regresja liniowa [Galton, 1877], model autoregresyjny [Yule, 1926], model autoregresyjny z ruchomą średnią [Kolmogorov, 1941; Wiener, 1949], regresja logistyczna [Cox, 1958], model autoregresyjny z wektorem błędu – VAR [Sims, 1980], model korekty błędu – ECM [Granger, 1981], model autoregresji z heteroskedastycznością warunkową – ARCH [Engle, 1982], uogólniony model autoregresji z heteroskedastycznością warunkową – GARCH [Bollerslev, 1986] lub model wektorowej autoregresji z wektorem błędu – VECM [Granger, Newbold, 1974].

Wskazane przykłady modeli ekonometrycznych istotnie różnią się od algorytmów i podejść wykorzystywanych w informatyce. Klasyczne[1] algorytmy IT działają według z góry określonych zasad, które są wprowadzane do algorytmu przez programistę na etapie tworzenia rozwiązania; każde działanie, operacja i decyzja uzyskane za pomocą rozwiązań IT są określone na podstawie instrukcji uwzględnionych w kodzie takiego rozwiązania [Kamiński i in., 2024]. W przypadku rozwiązań uczenia maszynowego działanie, operacje i decyzje, uzyskiwane za pomocą modelu, są wypadkową wzorców stosowanych w procesie uczenia z wykorzystaniem danych; model posiada cechy adaptacyjne zależne od danych zasilających proces uczenia takiego modelu. Oznacza to, że narzędzia ekonometryczne uczenia maszynowego były strukturalnie innymi podejściami od klasycznych programów lub algorytmów IT: w przypadku modeli ekonometrycznych pojedynczy model był uczony na podstawie wzorców danych, a nie bezpośrednio definiowany przez programistę (klasyczne IT).

[1] Wskazujemy na algorytmy, które są tworzone przez programistów i nie posiadają cech adaptacyjnych. Obecnie coraz więcej rozwiązań informatycznych posiada wbudowane mechanizmy modelu ekonometrycznych umożliwiających adaptacyjny charakter – te rozwiązania nie są rozumiane jako „klasyczne" IT.

Wskazane wcześniej przykłady modeli ekonometrycznych należały do klasy modeli parametrycznych, tj. modeli, w których zdefiniowana była postać relacji między zmiennymi objaśniającymi a zmienną objaśnianą [Hastie, Tibshirani, Friedman, 2009]. Na przykład model regresji liniowej odnosi się do sytuacji, w której twórca modelu antycypuje, że relacja między zmiennymi wprowadzanymi do modelu jest liniowa i zadana formułą:

Tabela 6.1. Proces generujący dane a model regresji liniowej

Założenie procesu generującego dane	Przyjęta postać modelu regresji liniowej
$y = X\beta + \varepsilon$ gdzie: y to wektor wartości zmiennej objaśnianej, X to macierz wartości zmiennych objaśniających, β to wektor parametrów prawdziwych procesu generującego dane, a ε to składnik losowy	$\hat{y} = X\hat{\beta}$ gdzie: $\hat{y}$ to wektor oszacowań wartości zmiennej objaśnianej, X to macierz wartości zmiennych objaśniających, a $\hat{\beta}$ to wektor oszacowanych parametrów modelu

Źródło: opracowanie własne.

W przypadku modeli parametrycznych, zakłada się „wprost" relację między zmiennymi objaśniającymi a zmienną objaśnianą; założenie to jest uchylone w przypadku modeli nieparametrycznych [Wasserman, 2006]. Modele tej klasy nie wymagają określenia struktury relacji między zmiennymi; w przeciwieństwie do modeli parametrycznych, w których struktura jest już zdefiniowana, modele nieparametryczne pozwalają na elastyczność w dopasowywaniu danych. Przykładami modeli nieparametrycznych są: metoda jądrowa [Parzen, 1962], drzewa decyzyjne [Quinlan, 1986], metoda najbliższych sąsiadów [Fix, 1985], regresja splajnów [Boor, 1978], metody regresji lokalnej [Loader, 2006] oraz modele funkcji łączonych [Hastie, Tibshirani, 1986]. Jeden z przykładów modeli nieparametrycznych stanowił model sieci neuronowych, wprowadzony w 1958 r. przez Franka Rosenblatta; był to model typu perceptron jako jedna z pierwszych sztucznych sieci neuronowych. Kolejne prace, m.in. Rumelharta, McClellanda i Parkera [1986], wprowadziły algorytm wstecznej propagacji błędu, umożliwiający efektywne trenowanie wielowarstwowych sztucznych sieci neuronowych.

Współczesnym nurtem związanym z modelami ekonometrycznymi jest rozwój modeli opartych na sieciach neuronowych – ich przykładem są generatywne sieci neuronowe, umożliwiające tworzenie nowych danych na podstawie danych historycznych, na których taka sieć była trenowana [Goodfellow i in., 2014]. Innowacją w zastosowaniu tych podejść jest wyjście poza czysto liczbowy przedmiot generowania prognoz, tak jak w przypadku wskazanych wcześniej podejść historycznych.

Modele generatywnej sztucznej inteligencji umożliwiają tworzenie nowych treści, takich jak teksty, obrazy oraz inne formaty danych [Shanmugamani, 2018]. Obecnie występuje zróżnicowanie zastosowań, które dzisiaj można wpisać pod nazwę generatywnych narzędzi sztucznej inteligencji: są to zarówno generatory tekstu, takie jak ChatGPT, BERT, jak też mechanizmy do tworzenia kodu źródłowego np. GitHub Copilot lub generowania obrazu jak DALL-E. Szczególnym przypadkiem modeli generatywnych, których początek można wskazywać na lata 80., są duże modele językowe (*large language models*, LLM).

6.1. Duże modele językowe

LLM są modelami przetwarzania danych tekstowych, tzw. przetwarzania języka naturalnego, bazujące w dużym stopniu na strukturze transformera[2] [Vaswani i in., 2017]. Modele LLM są trenowane na szerokich zbiorach danych tekstowych, np. anglojęzyczna Wikipedia zawiera 7 mln artykułów [Rothman, 2021]. Zastosowanie tych modeli obejmuje obszary związane z przetwarzaniem tekstu, m.in. tworzenie nowych treści, rozpoznawanie mowy, generowanie streszczeń czy też przygotowanie określonych szablonów tekstowych, np. formatowanie danego tekstu na podstawie podanego wzoru [Vajjala, Majumder, Gupta, Surana, 2020]. Modele LLM dzieli się na małe, średnie i duże. Podział ten definiowany jest liczbą parametrów modelu: małe posiadają niespełna kilka miliardów parametrów, średnie nawet do kilkudziesięciu miliardów i duże opisywane są w setkach miliardów parametrów. Klasyfikację modeli LLM ze względu na ich wielkość przedstawiono w tabeli 6.2.

Tabela 6.2. Klasyfikacja modeli LLM ze względu na ich wielkość

Rozmiar	Liczba parametrów	Przykład	Wymagania sprzętowe
Małe	do kilku mld	DistilBERT, Phi 2, GPT-Neo	komputer osobisty, GPU
Średnie	10–50 mld	Falcon, BloombergGPT	małe serwery
Duże	od 50 do setek mld	Megatron-Turing NLG, GPT-4, Minerva	duże serwery

Źródło: opracowanie własne.

[2] Jest to typ sieci neuronowej, która wykrywa kontekst sekwencyjnych danych i na jego podstawie generuje odpowiedź. Transformer składa się z enkodera (*encoder*), który ma za zadanie zrozumieć dane wejściowe, oraz dekodera (*decoder*), którego zadaniem jest wygenerowanie danych wyjściowych na podstawie danych uzyskanych w enkoderze.

Podczas trenowania modeli LLM mogą wystąpić zjawiska niedostatecznego albo nadmiernego dopasowania danych. W przypadku dużych modeli językowych zjawiska te spowodowane są rozmiarem (tj. liczbą parametrów modelu) oraz długością trenowania modelu (tj. liczbą iteracji na zbiorze treningowym). Oceniając jakość modelu LLM na innym zbiorze danych niż zbiór danych treningowych, można się spodziewać, że zależność między poziomem błędu a iteracjami będzie miała kształt litery U; na początku trenowania błąd modelu LLM spada, przy czym od pewnego momentu następuje przeuczenie modelu i błąd zaczyna ponownie rosnąć. Kiedy znajdzie się w punkcie będącym minimum błędu, zwiększenie liczby parametrów bądź iteracji będzie równoznaczne z wystąpieniem zjawiska nadmiernego dopasowania do danych testowych oraz zmniejszeniem dopasowania do danych populacji ogólnej. W rzeczywistości natomiast, zwiększając te wartości, można przekroczyć pewne maksimum, do którego wartość miary błędu predykcji rośnie. Następnie ta wartość zaczyna spadać poniżej punktu wcześniej określanego jako optymalny [Nakkiran i in., 2021], będącego minimum wcześniejszego wykresu. W konsekwencji nowy wykres byłby w kształcie odwróconej litery N. Zjawisko to nosi nazwę podwójnego spadku (*double descent*).

Moc obliczeniowa potrzebna do trenowania dużych modeli językowych w dużym stopniu wynika z liczby parametrów, jakie posiadają te modele [Kaplan i in., 2020]. Modele LLM, które kategoryzuje się jako małe, można trenować nawet na komputerach osobistych. W przypadku średnich modeli LLM, które zawierają już kilkadziesiąt miliardów parametrów, trenowanie powinno odbywać się na mniejszych serwerach. Trenowanie dużych modeli LLM wymaga natomiast ogromnych nakładów finansowych, mocy obliczeniowej, którą można uzyskać tylko w dużych centrach danych. Na przykład wytrenowanie modelu GPT-3 kosztowało około 12 mln USD [Jordan, 2020].

Duże modele LLM wykorzystuje się w różnych obszarach, kontekstach, problemach, takich jak: tłumaczenie tekstu, generowanie kodu źródłowego, odpowiadanie na pytania, podsumowanie tekstu, analiza semantyczna [Zharovskikh, 2023]. Modele klasy LLM wykorzystywane są m.in. w finansach do wykrywania oszustw lub analizy dokumentów finansowych, w prawie do sporządzania dokumentów oraz znajdowania luk prawnych, a w medycynie do wykrywania chorób i tworzenia planów leczenia.

W zależności od zadania, do którego modele będą wykorzystywane, oceniane są one przy pomocy różnych metryk ewaluacji, tj. miar oceny działania danego modelu. W kolejnej sekcji tekstu przedstawiono standardowe miary oceny działania modeli LLM.

6.2. Pomiar jakości działania LLM

Nie istnieje jeden sposób pomiaru jakości działania modeli LLM [Chang i in., 2024]. Miary te definiowane są na podstawie obszaru działania modelu; inne miary oceny jakości mogą być wykorzystywane np. dla streszczeń tekstu, a inne dla generowania kodu źródłowego bądź tłumaczeń. Poniżej przedstawiono zestawienie najpopularniejszych miar oceny jakości działania modeli LLM.

6.3. Recall-Oriented Understudy for Gisting Evaluation (ROUGE)

Jednym z przykładów miar jakości działania modeli LLM jest miara Recall-Oriented Understudy for Gisting Evaluation, zwana również krócej ROUGE [Lin, Chin-Yew, 2004]. Jest to metryka służąca do oceny jakości streszczeń lub tłumaczeń wytworzonych przez model LLM względem tekstu utworzonego przez człowieka [Lin, Chin-Yew, 2004]. Wartości, jakie miara ROUGE może przyjmować, obejmują zakres: 0–1, gdzie 0 wskazuje na całkowity brak podobieństwa tekstu, a 1 – na identyczność względem tekstu referencyjnego.

ROUGE jest w istocie zbiorem miar, różniących się od siebie rozmiarem n-gramów[3], które będą do siebie porównywane. Na przykład w przypadku ROUGE-2 porównuje się bigramy (pary wyrazów). Porównując zdania: „Jan jest studentem ekonomii" i „Kuba jest studentem fizyki", najpierw należy podzielić zdania na bigramy. W pierwszym przypadku są to: „Jan jest", „jest studentem", „studentem ekonomii". Natomiast w drugim są to: „Kuba jest", „jest studentem", „studentem fizyki". Następnie liczy się powtarzające się w obu przypadkach pary i dzieli przez liczbę bigramów występujących w tekście referencyjnym, w tym przypadku wartość ta wynosi 3. ROUGE-2 przyjmuje więc wartość 1/3 = 0,33. Inne miary ROUGE to ROUGE-1 i ROUGE-L, gdzie dzielimy najdłuższy wspólny szereg słów przez całkowitą liczbę słów w tekście referencyjnym.

6.4. Perpleksja – *perplexity*

Innym przykładem miary oceny jakości działania modeli LLM jest miara perpleksji (*perplexity*). Jest ona oparta na funkcji wiarygodności, umożliwiającej określenie prawdopodobieństwa pojawiania się danego wyrazu w kontekście danego zdania. Im

[3] Jako *n*-gram rozumie się *n*-elementowy szereg słów ustawionych w danej kolejności, występujących w określonej formie odmiany.

większa wartość, tym mniejsze prawdopodobieństwo pojawienia się danego tekstu (tzn. tekst jest mniej prawdopodobny).

$$\log(PPL) = -\frac{1}{N}\sum_{i=1}^{N}\log(P(t_i \mid c_{t_i})),$$

gdzie: *PPL* to miara perpleksji, *N* oznacza rozmiar zbioru testowego, *P* reprezentuje prawdopodobieństwo, t_i oznacza token „i", dla którego obliczana jest wartość miary pod warunkiem poprzedzających go tokenów c_{t_i}.

Poprzez token rozumie się formalną (matematyczną) reprezentację danego wyrazu[4].

Poniżej przedstawiono przykład działania modelu LLM przy różnych stopniach jego wytrenowania. Demonstruje on, jak model językowy GPT-2 odpowiada na identyczne zapytanie przy trzech różnych poziomach wytrenowania. Doszkalanie modelu zostanie dokładniej omówione w kolejnej sekcji tekstu.

Pierwszym z trzech jest model surowy, którego parametry przyjmują losowe wartości (model bez wytrenowania). Tekst generowany przez taki model składa się z losowych tokenów, tj. słów lub znaków. Jego metryka, średnia perpleksja, jest stosunkowo wysoka, co oznacza, że niepewność odnośnie do generowanego tekstu jest wysoka.

Następnie to samo zapytanie zostało przesłane do wytrenowanego, ogólnego modelu. Wygenerowany tekst składa się już z poprawnych zdań, jednak pozbawiony jest spójności tematycznej oraz głębszego zrozumienia kontekstu. Miara perpleksji, czyli miara niepewności, jest znacząco niższa, co przekłada się na tekst bardziej zgodny z oczekiwaniami opartymi na danych wejściowych.

Tabela 6.3. Przykładowe treści wygenerowane przez model GPT-2 dla modelu surowego, ogólnego oraz wyspecjalizowanego

Model	Wyjście	Średnia perpleksja
Surowy model	**Economics is** afterlifeç gayFine Alas Alas Alasjadjadjadjadjadjad Madurojadjadjad	17 406
Wytrenowany ogólny model	**Economics is** a very important field for the study of economics	3522
Wyspecjalizowany model	**Economics is** a rich and diverse region. Inclusion of diverse economic and social issues in a diverse region is essential to succeed	854

Źródło: opracowanie własne.

[4] Dla przykładu słowo *reading* w zależności od przyjętej reprezentacji może oznaczać jeden lub kilka tokenów ([„read", „ing"] lub [„reading"]). Słowniki tokenów różnią się pomiędzy modelami językowymi, stąd różnice w reprezentacji słów lub ich części.

Ostatnie zapytanie zostało skierowane do modelu wyspecjalizowanego i doszkolonego (*fine-tuned*). W tym przypadku został on wytrenowany na ekonomicznej bazie danych. Tekst wygenerowany przez taki model jest bardziej złożony i spójny, a niepewność w zakresie pojawiających się słów w zdaniu – mniejsza względem całego kontekstu.

6.5. *Cross entropy loss*

Jednym ze sposobów oceny modeli oraz monitorowania przebiegu ich treningu jest wykorzystanie funkcji straty (*loss function*). Funkcje te służą do obliczania wartości, zwanej stratą walidacyjną (*validation loss*), która mierzy utratę skuteczności w generowaniu odpowiedzi względem tekstów walidacyjnych. Mogą się one różnić w zależności od modelu czy zadania, do którego model jest trenowany. Jedna z tych funkcji to strata entropii krzyżowej (*cross entropy loss*). Funkcja wykorzystywana do obliczenia wartości straty ma postać:

$$L = \frac{1}{N}\sum_{i=1}^{N} -\log \hat{y}^{(i)}_{x_{i+1}},$$

gdzie: N to rozmiar zbioru testowego, a L to wartość funkcji straty.

Funkcja ta dla każdego i liczy entropię krzyżową pomiędzy przewidzianym rozkładem prawdopodobieństwa $\hat{y}^{(i)}$ a kolejnym słowem zaobserwowanym w tekście referencyjnym. Następnie liczona jest średnia entropia dla całego zbioru walidacyjnego, ale funkcja ta może także zostać użyta do obliczenia straty w zbiorze testowym czy treningowym.

Obliczenie wartości funkcji straty wygląda w następujący sposób: na początku niezbędny jest tekst, do którego model będzie porównywany. Przyjmijmy, że jest nim zdanie: „Jan gra w piłkę nożną". Następnie generowany jest tekst. Podczas generowania powstają rozkłady prawdopodobieństwa dla każdego kolejnego słowa, z pominięciem pierwszego, które zostało uwzględnione w komendzie. Zakładając, że wygenerowany tekst brzmi: „Jan grał w piłkę ręczną", wartość funkcji straty jest liczona, jeśli wygenerowane słowo nie jest tym samym co słowo w tekście referencyjnym i przyjmuje ona wtedy wartość ujemnego logarytmu prawdopodobieństwa wystąpienia tego słowa. Jeśli słowo to odpowiada słowu na tej samej pozycji w tekście referencyjnym, strata jest równa 0. W wygenerowanym zdaniu słowa, dla których wartość funkcji straty będzie różna od 0, to „Jan" i „piłkę", ponieważ następne słowa nie odpowiadają tym referencyjnym. Na koniec liczona jest średnia ze wszystkich wartości i otrzymany wynik stanowi wartość funkcji straty dla całego tekstu.

6.6. Doszkalanie LLM

Doszkalanie modelu (*fine-tuning*) to proces, który polega na dostosowaniu już wcześniej wytrenowanego modelu do specyficznego kontekstu dziedzinowego [Devlin i in., 2019], w którym to kontekście model będzie wykorzystywany (np. wcześniej wskazywane przykłady z prawa lub medycyny). W przypadku LLM doszkalanie modelu polega na zmianie części jego parametrów w celu dostosowania wydajności do konkretnego kontekstu bez konieczności trenowania całego modelu od początku [Khandare, 2023]. Jest to szczególnie użyteczne, gdy zadanie, w ramach którego model ma być wykorzystywany, różni się od oryginalnego zadania, do którego został przeznaczony. Istnieje natomiast kilka rodzajów technik doszkalania modeli, wśród których można wyróżnić:

- **całkowite doszkalanie modelu** – polegające na trenowaniu całego modelu na podstawie nowego zbioru danych; taki typ doszkalania wymaga znacznych nakładów mocy obliczeniowej oraz wystarczająco dużego zbioru danych, na podstawie którego trenowane będą wszystkie warstwy modelu;
- **selektywne doszkalanie modelu** – technika polegająca na dostrojeniu tylko wybranych warstw modelu, zwykle górnych lub dolnych (inaczej: początkowych bądź końcowych); wymaga znacznie mniejszej mocy obliczeniowej w porównaniu z całkowitym doszkalaniem modelu i jest używana w przypadku, gdy zadanie docelowe jest zbieżne z pierwotnym zadaniem ogólnego modelu;
- **efektywne doszkalanie modelu** (*parameter-efficient fine-tuning*, PEFT) – charakteryzujące się dodaniem małej warstwy nowych parametrów w celu dostosowania modelu pod konkretne zadanie; pozostałe parametry pozostają niezmienne, co zmniejsza potrzebną moc obliczeniową;
- **hybrydowe doszkalanie modelu** – polegające na połączeniu różnych podejść doszkalania w celu zachowania części oryginalnych cech modelu z jednoczesnym dodaniem nowych; proces doszkalania dzieli się tutaj na kilka etapów, w których raz trenowana jest większa część parametrów, a raz mniejsza; proces ten pozwala na dostrojenie modelu do bardzo specyficznych zadań, nie pozbawiając go jednocześnie jego podstawowych możliwości.

6.7. Przykład doszkalania LLM

W ramach badania przeprowadzony został eksperyment numeryczny, polegający na wykorzystaniu modelu BART opracowanego przez firmę Facebook [Lewis i in., 2019].

Celem eksperymentu było sprawdzenie, w jaki sposób doszkalanie modelu LLM na wyspecjalizowanym zbiorze danych (artykuły medyczne) wpływa na jakość generowanych treści na innym ogólnym zbiorze danych (tu: zawierającym artykuły z serwisu internetowego telewizji CNN).

Dane medyczne zawierały 20 tys. obserwacji w zbiorze treningowym, użytych do dotrenowania modelu i dodatkowych 1000 w zbiorze walidacyjnym. Składały się one z abstraktów biomedycznych i odpowiadających im streszczeń.

Tabela 6.4. Przykład obserwacji ze zbioru danych medycznych

Artykuł	Streszczenie
„Cardiovascular disease is the leading cause of death globally. While pharmacological advancements have improved the morbidity and mortality associated with cardiovascular disease, non-adherence to prescribed treatment remains a significant barrier to improved..."	"Medication adherence in cardiovascular medicine"

Źródło: opracowanie własne.

Dane nazwane jako CNN zawierały 1000 obserwacji użytych do walidacji modelu. Zbiór składał się z artykułów ukazanych w CNN i odpowiadających im streszczeń.

Tabela 6.5. Przykład obserwacji ze zbioru artykułów CNN

Artykuł	Streszczenie
„The Palestinian Authority officially became the 123 rd member of the International Criminal Court on Wednesday, a step that gives the court jurisdiction over alleged crimes in Palestinian territories. The formal accession was marked with a ceremony at The Hague, in the Netherlands..."	"Membership gives the ICC jurisdiction over alleged crimes committed in Palestinian territories since last June..."

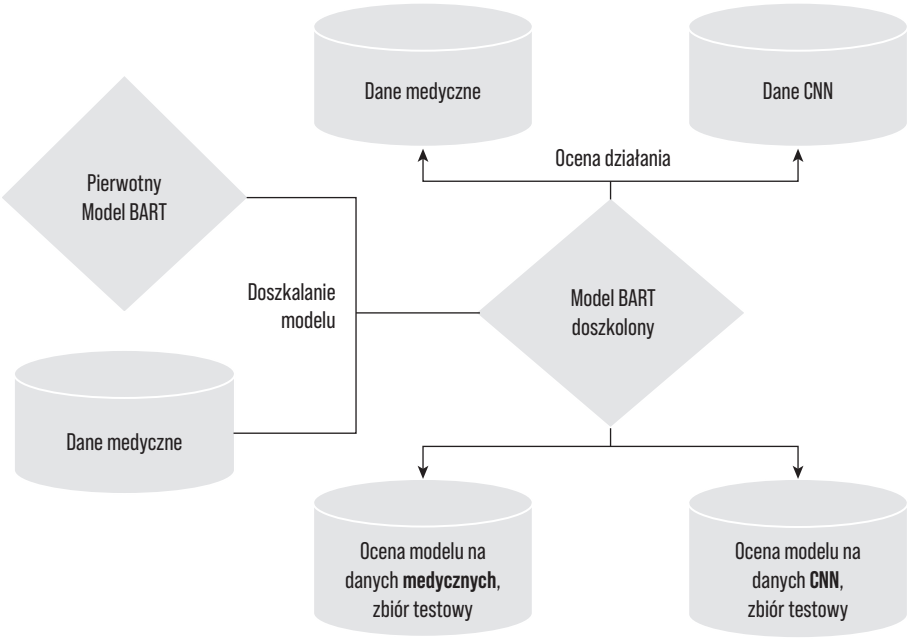
Źródło: opracowanie własne.

Schemat przeprowadzonego eksperymentu numerycznego przedstawiono na rysunku 6.1.

W pracy zastosowano efektywne doszkalanie modelu – PEFT, przeprowadzone w sześciu iteracjach (*epoch*). Liczba parametrów modelu użytego w eksperymencie wynosi 139 641 600. Następnie zostało dodane 221 184 parametry, co stanowi 0,16% wszystkich parametrów modelu początkowego.

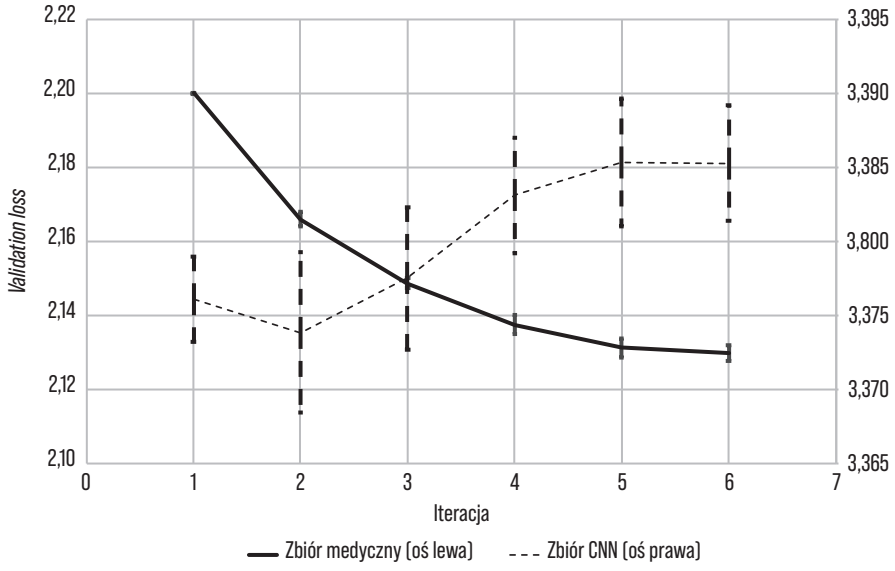
Średnie wartości błędów na zbiorze walidacyjnym, odpowiednio dla danych medycznych oraz danych CNN, uzyskane w oparciu na modelu doszkalanym na zbiorze medycznym, przedstawiono na rysunku 6.2 (wartości liczbowe w załączniku A).

Rysunek 6.1. Schemat eksperymentu numerycznego



Źródło: opracowanie własne.

Rysunek 6.2. Schemat eksperymentu numerycznego



Źródło: opracowanie własne.

Wraz z każdą iteracją doszkalania uzyskiwano statystycznie lepszą jakość modelu na danych medycznych (linia ciągła na rysunku 6.2). Oznacza to, że model generował lepszej jakości tekst w kontekście medycznym, co jest zjawiskiem oczekiwanym: doszkalając model w oparciu na specyficznych danych medycznych, oczekujemy, że będzie on lepiej działał w tym konkretnym kontekście dziedzinowym.

Interesujący jest jednak profil jakości działania modelu dla danych zbioru CNN (linia przerywana na rysunku 6.2). Obszar ten jest w dużej mierze niezależny od danych medycznych, obejmując kwestie społeczne, socjologiczne, prawne i ekonomiczne, które co do zasady są niezależnym obszarem względem medycyny[5]. Z każdą iteracją doszkalania modelu BART, w oparciu na danych medycznych, ten sam model wykorzystywany do analizy danych ze zbioru CNN uzyskiwał coraz wyższy błąd *validation loss*. Obserwowany jest statystyczny spadek jakości działania modelu w oparciu na danych CNN.

Zidentyfikowana w pracy anomalia to szerzej znane w literaturze katastrofalne zapominanie (*catastrophic forgetting*), związane z sytuacją, w której model poprzez doszkalanie w jednym obszarze wykorzystania (tu: dane medyczne) traci dokładność działania w innym niezależnym obszarze (tu: dane CNN) [Kirkpatrick i in., 2017]. Zjawisko to jest obecne m.in. w przypadku architektur modeli opartych na sieciach neuronowych. Za główną przyczynę wspomnianej anomalii wskazuje się mechanizm dostosowywania wag modelu. Podczas doszkalania modelu w danym obszarze potencjalnie wszystkie wagi modelu mogą ulegać nieznacznym modyfikacjom, co niekorzystnie wpływa na jakość działania modelu na innych obszarach wykorzystania.

Podsumowanie

Rozwój ekonometrii – jak zaznaczono we wstępie, dziedziny leżącej pomiędzy matematyką, statystyką oraz informatyką – wiązał się na początku z paradygmatem modelowania parametrycznego (tj. modelami o zadanej postaci funkcyjnej), a później z przejściem na modele nieparametryczne. Przykładem klasy modeli nieparametrycznych są sztuczne sieci neuronowe. Początkowe wykorzystanie tych modeli, skutkujące poprawą względem wcześniejszych podejść w obszarze jakości generowanych pro-

[5] Autorzy dostrzegają związki między kwestiami społecznymi, socjologicznymi, prawnymi lub wskazaniami ekonomicznymi a medycyną; niemniej dane zbioru CNN odnoszą się do kwestii medycznych wyłącznie powierzchownie, nie traktując tego jako głównego obszaru zainteresowań. Dla uproszczenia wywodu przyjęto, że zagadnienia objęte w danych CNN są niezależne od zagadnień objętych danymi medycznymi.

gnoz z modeli, przyniosło dalszy rozwój i zaawansowanie wykorzystywanych podejść. Obecnie rozwój tego obszaru widać na przykładzie popularnych, dostępnych dla użytkowników nietechnicznych, dużych modeli językowych.

Ponadto obserwujemy coraz większą popularyzację wykorzystania modeli LLM w wymiarze zarówno horyzontalnym: różne konteksty dziedzinowe, jak również wertykalnym: szczegółowe i specjalistyczne zadania do wykonywania. Jednocześnie budowa modelu LLM od podstaw jest z reguły ekonomicznie nieuzasadniona z punktu widzenia pojedynczego przedsiębiorstwa.

Taki stan rzeczy powoduje, że przedsiębiorstwa coraz częściej wykorzystują modele językowe – wytrenowane i udostępnione przez duże firmy technologiczne. W celu dostosowania sposobu działania tych modeli z charakteru ogólnego na wyspecjalizowany, dziedzinowy wypracowano szereg metod ich doszkalania.

Proces doszkalania, w zależności od rodzaju, może dotyczyć wszystkich lub wybranych parametrów sieci neuronowej. Co więcej, występują podejścia (takie zastosowano w pracy), w których – zamiast modyfikacji istniejących parametrów – do sieci neuronowej wprowadza się dodatkowe warstwy i to na nich wykonywana jest modyfikacja.

Doszkalanie modeli powoduje poprawę dokładności działania modelu w dziedzinie, w której był doszkalany. W pracy wskazano jednak scenariusz, w którym obok poprawy dokładności działania modelu w konkretnej dziedzinie, w jakiej wykonywane było doszkalanie, zaobserwowano również pogorszenie oceny działania modeli w innej dziedzinie. Omawiany proces w literaturze nazywa się katastrofalnym zapominaniem i związany jest z naruszeniem wartości wag modelu, odpowiadających za obszar, w którym model nie był doszkalany (w przykładzie były to dane CNN).

Obserwacja poczyniona w pracy dotyczy faktu, że przejście z modeli ogólnych na modele domenowo wyspecjalizowane, poprzez procedurę doszkalania modelu, może powodować nieplanowane pogorszenie działania tego modelu w obszarze niezależnym. Oznaczać to może, że wraz z postępującym doszkalaniem modelu w konkretnym obszarze zastosowań, model ten powinien być również wyłączany z pozostałych niezależnych zastosowań.

Załącznik

Tabela Z6.1. Strata walidacyjna modelu dotrenowanego na danych medycznych i ocenionego na danych medycznych

Iteracja	Średnia	Odchylenie standardowe	Przedział ufności
1	2,20021	0,00005	0,00003
2	2,16601	0,00299	0,00185
3	2,14860	0,00220	0,00137
4	2,13749	0,00410	0,00254
5	2,13114	0,00419	0,00260
6	2,12980	0,00336	0,00208

Źródło: opracowanie własne.

Tabela Z6.2. Strata walidacyjna modelu dotrenowanego na danych medycznych i ocenionego na artykułach z CNN

Iteracja	Średnia	Odchylenie standardowe	Przedział ufności
1	3,37608	0,00459	0,00285
2	3,37384	0,00874	0,00542
3	3,37749	0,00773	0,00479
4	3,38311	0,00634	0,00393
5	3,38535	0,00701	0,00435
6	3,38530	0,00631	0,00391

Źródło: opracowanie własne.

Bibliografia

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, 31(3), s. 307–327.

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., Ye, W., Zhang, Y., Chang, Y., Yu, P.S., Yang, Q., Xie, X. (2024). A Survey on Evaluation of Large Language Models, *ACM Transactions on Intelligent Systems and Technology*, 15(3), s. 1–45.

Cox, D.R. (1958). The Regression Analysis of Binary Sequences, *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 20(2), s. 215–232.

De Boor, C. (1978). *A Practical Guide to Splines*. New York: Springer.

Devlin, J., Chang, M., Lee, K., Toutanova, K. (2019). BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. W: *North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (s. 4171–4186). Minneapolis, Minnesota: Association for Computational Linguistics.

Engle, R.F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica: Journal of the Econometric Society*, 50(4), s. 987–1007.

Fix, E. (1985). *Discriminatory Analysis: Nonparametric Discrimination, Consistency Properties* (vol. 1). Texas: USAF School of Aviation Medicine.

Frisch, R. (1926). *Sur un problème d'économie pure*. Grøndahl & søns boktrykkeri.

Galton, F. (1877). *Typical Laws of Heredity*. William Clowes and Sons.

Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y. (2014). Generative Adversarial Nets. W: *Advances in Neural Information Processing Systems* (vol. 3, s. 2672–2680). DOI: 10.1007/978-3-658-40442-0_9.

Granger, C.W. (1981). Some Properties of Time Series Data and Their Use in Econometric Model Specification, *Journal of Econometrics*, 16(1), s. 121–130.

Granger, C.W., Newbold, P. (1974). Spurious Regressions in Econometrics, *Journal of Econometrics*, 2(2), s. 111–120.

Hastie, T., Tibshirani, R. (1985). Generalized Additive Models; Some Applications. W: *Generalized Linear Models: Lecture Notes in Statistics* (vol. 32, s. 66–81), R. Gilchrist, B. Francis, J. Whittaker (Eds). New York: Springer.

Hastie, T., Tibshirani, R., Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). New York: Springer.

Jordan, J. (2020). *The Cost of Training GPT-3*, https://towardsdatascience.com/the-cost-of-training-gpt-3-7-m-8b9b1c472fbf (dostęp: 30.04.2024).

Kamiński, B., Król-Całkowska, J., Raulinajtys-Grzybek, M., Więckowska, B., Trzebiński, W., Byszek, K., Jaśkowiak, D., Kaszyński, D. (2024). *Sztuczna inteligencja w zdrowiu. Bezpieczeństwo prawne i wykorzystanie w Polsce*. SGH Think Tank dla ochrony zdrowia. DOI: 10.33119/SZIWZBPIWWP-2024-1-81.

Kaplan, J., McCandlish, S., Henighan, T., Brown, T.B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., Amodei, D. (2020). *Scaling Laws for Neural Language Models*. DOI: 10.48550/arXiv.2001.08361.

Khandare, S.S. (2023). *Mastering Large Language Models: Advanced Techniques, Applications, Cutting-Edge Methods, and Top LLMs* (English ed.). Independently published.

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A.A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., Hassabis, D. (2017). Overcoming Catastrophic Forgetting in Neural Networks, *Proceedings of the National Academy of Sciences*, 114(13), s. 3521–3526.

Kolmogorov, A. (1941). Interpolation and Extrapolation of Stationary Random Sequences. *Izvestiya the Academy of Sciences of the USSR*, 5, s. 3–14.

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., Zettlemoyer, L. (2019). *BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension*. DOI: 10.48550/arXiv.1910.13461.

Lin, C.Y. (2004, July). Rouge: A Package for Automatic Evaluation of Summaries. W: *Text Summarization Branches Out* (s. 74–81). Barcelona: Association for Computational Linguistics.

Loader, C. (2006). *Local Regression and Likelihood*. New York: Springer.

Nakkiran, P., Kaplun, G., Bansal, Y., Yang, T., Barak, B., Sutskever, I. (2021). Deep Double Ddescent: Where Bigger Models and More Data Hurt, *Journal of Statistical Mechanics: Theory and Experiment*, 12, 124003.

Parzen, E. (1962). On Estimation of a Probability Density Function and Mode, *The Annals of Mathematical Statistics*, 33(3), s. 1065–1076.

Quinlan, J.R. (1986). Induction of Decision Trees, *Machine Learning*, 1, s. 81–106.

Rothman, D. (2021). *Transformers for Natural Language Processing: Build Innovative Deep Neural Network Architectures for NLP with Python, PyTorch, TensorFlow, BERT, RoBERTa, and More*. Birmingham: Packt Publishing.

Rouge, L.C. (2004). A Package for Automatic Evaluation of Summaries. W: *Text Summarization Branches Out* (s. 74–81). Barcelona: Association for Computational Linguistics.

Shanmugamani, R. (2018). *Deep Learning for Computer Vision: Expert Techniques to Train Advanced Neural Networks Using TensorFlow and Keras*. Birmingham: Packt Publishing.

Sims, C.A. (1980). Macroeconomics and Reality, *Econometrica: Journal of the Econometric Society*, 48(1), s. 1–48.

Vajjala, S., Majumder, B., Gupta, A., Surana, H. (2020). *Practical Natural Language Processing: a Comprehensive Guide to Building Real-World NLP Systems*. Sebastopol: O'Reilly Media.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I. (2017). *Attention Is All You Need*, *Neural Information Processing Systems*, 30, s. 5998–6008.

Wasserman, L. (2006). *All of Nonparametric Statistics*. New York: Springer.

Wiener, N. (1949). *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications*. The MIT Press. DOI: 10.7551/mitpress/2946.001.0001.

Winston, P.H. (1977). *Artificial Intelligence*. Massachusetts: Addison-Wesley.

Yule, G.U. (1926). Why do We Sometimes Get Nonsense Correlations Between Time-Series?, *Journal of the Royal Statistical Society*, 89(1), s. 1–63.

Zharovskikh, A. (2023). *Best Applications of Large Language Models*, https://indatalabs.com/blog/large-language-model-apps (dostęp: 30.04.2024).

7.0

ZBIORY DANYCH I MODELE JĘZYKOWE STOSOWANE W JĘZYKU POLSKIM DO IDENTYFIKACJI STWIERDZEŃ WYMAGAJĄCYCH WERYFIKACJI

Krzysztof Węcel
Uniwersytet Ekonomiczny w Poznaniu
Marcin Sawiński
Uniwersytet Ekonomiczny w Poznaniu
Ewelina Księżniak
Uniwersytet Ekonomiczny w Poznaniu
Milena Stróżyna
Uniwersytet Ekonomiczny w Poznaniu

Streszczenie

W rozdziale opisano metodykę tworzenia zbioru tekstów w języku polskim oraz trenowania modeli do wsparcia procesu weryfikacji informacji pod kątem ich prawdziwości i dokładności. Opracowany został protokół oznaczania tekstów jako „check-worthy". Posłużył on do opracowania własnego zbioru w języku polskim, wykorzystanego następnie do trenowania i ewaluacji wybranych modeli z różnymi wariantami zbiorów danych, uwzględniających zasoby w języku polskim oraz mieszane. Przeprowadzone badania wskazują, że modele przeznaczone dla języka polskiego oraz modele wielojęzyczne miały zbliżoną dokładność. Ważną rolę odgrywa sposób doboru przykładów do treningu. Uzasadnione jest wykorzystywanie modeli wielojęzycznych, umożliwiających trenowanie na większej liczbie zasobów obcojęzycznych, i dotrenowanie

na wysokiej jakości przykładach w języku polskim. Wynikiem badań jest wysokiej jakości zbiór sklasyfikowanych tekstów w języku polskim do trenowania modeli oraz najlepszy model dla języka polskiego służący do określania, czy dane stwierdzenie wymaga weryfikacji.

Słowa kluczowe: jakość informacji, wiarygodność informacji, modele językowe, fact-checking

Wprowadzenie

Jednym z kluczowych wymogów dla skutecznego zarządzania w firmie jest dostęp do prawdziwej, wiarygodnej, aktualnej i relewantnej informacji, którą określamy również jako informację wysokiej jakości. We współczesnym świecie, mimo nadmiaru informacji, obserwujemy coraz większy problem z dostępem do jakościowej informacji. Z jednej strony jest to wynikiem demokratyzacji dostępu do tworzenia treści, gdzie każdy może opublikować dowolną informację. Z różnych pobudek może być ona jednak nieprawdziwa, a celem jej tworzenia może okazać się chęć uzyskania rozgłosu, osiągnięcia konkretnych korzyści czy wywołania chaosu wśród konkurentów. Z drugiej strony mamy do czynienia z intensywnym rozwojem metod sztucznej inteligencji, wykorzystanej do automatycznego tworzenia treści. W zależności od intencji one również mogą być nieprawdziwe. Znane jest także zjawisko spontanicznego tworzenia nieprawdziwych treści przez sztuczną inteligencję, określane jako halucynacje.

Zalew fałszywych informacji w Internecie stanowi wyzwanie zarówno dla działania firm (np. nieuczciwe opinie o podmiocie), podejmowania decyzji przez konsumentów (np. nadmiernie optymistyczne opisy produktów), jak i dla bezpieczeństwa publicznego (np. ruchy antyszczepionkowe). Coraz bardziej zwraca się uwagę na systemową walkę z tym zjawiskiem, czego dowodem są niedawne regulacje unijne (Digital Services Act, DSA), które wprowadzają szereg obowiązków w zakresie przeciwdziałania rozprzestrzenianiu się dezinformacji. Na przykład bardzo duże platformy internetowe i wyszukiwarki (gdy liczba użytkowników przekracza 10% ludności UE) są zobowiązywane do usuwania dezinformujących treści.

W zapobieganiu rozprzestrzeniania dezinformacji istotną rolę odgrywają tzw. agencje fact-checkingowe, które zawodowo zajmują się monitorowaniem i weryfikacją prawdziwości informacji. Fact-checking oznacza rzetelny proces weryfikacji informacji pod kątem ich prawdziwości i dokładności. Obejmuje on krytyczną analizę poszczególnych stwierdzeń lub danych prezentowanych w różnych formach i mediach. Ele-

mentem tej oceny jest wcześniejszy wybór, z pojedynczego dokumentu lub całego zbioru, stwierdzeń wymagających weryfikacji (*check-worthy*). Są to takie stwierdzenia, które faktycznie powinny zostać poddane procesowi fact-checkingu ze względu na swoje potencjalne oddziaływanie na decyzje innych oraz prawdopodobieństwo tego, że są to stwierdzenia fałszywe.

7.1. Zagadnienie klasyfikacji treści tekstowych

Fact-checking pierwotnie rozumiany był jako potwierdzenie dokładności informacji, zamieszczonych w artykule przed samą jego publikacją. A zatem był to proces wewnętrzny, związany z odpowiedzialnością dziennikarzy za swoją pracę. Z czasem znaczenie tego pojęcia zostało rozszerzone o analizę wiadomości już po ich opublikowaniu, przede wszystkim w Internecie, wykonywaną przez osoby, które nie były autorami pierwotnego tekstu [Cazalens i in., 2018].

Proces fact-checkingu składa się z kilku etapów. Pierwszym jest wybór stwierdzeń do weryfikacji. Po nim następują kontekstualizacja i analiza, konfrontacja z zewnętrznymi bazami danych oraz wiedzą ekspertów. Całość kończy przygotowanie opracowania z przeprowadzonych badań wraz z podjęciem decyzji co do oceny informacji, np. prawda, fałsz, manipulacja, brak kontekstu [Micallef i in., 2022].

Głównym wyzwaniem obecnie jest to, że większość zadań w ramach powyższego procesu wykonywana jest manualnie. Ze względu na ilość informacji, które wymagają weryfikacji, wskazana jest jego automatyzacja. Jako lukę badawczą należy zatem wskazać brak narzędzi wspierających fact-checkerów, szczególnie w języku polskim, które ułatwiłyby im pracę, zwiększyły dokładność, przyspieszyły podejmowanie decyzji, a także pozwoliłyby na skuteczniejsze wykrywanie fake newsów.

Etapem procesu, któremu poświęcono ten rozdział, jest identyfikacja stwierdzeń do weryfikacji. Należy ona do ogólniejszego zagadnienia klasyfikacji treści tekstowych, tj. przypisywania określonej etykiety do podanego tekstu. Jeśli weźmie się przykład z analizy wydźwięku (*sentiment analysis*), tekst może być klasyfikowany jako „pozytywny", „negatywny" lub „neutralny". W naszym konkretnym przypadku będą to dwie klasy (klasyfikacja binarna): „wymagające weryfikacji" oraz „nieistotne".

Klasyfikacja tekstu odbywa się z wykorzystaniem odpowiedniego modelu, który jest wcześniej uczony maszynowo. Polega to na pokazywaniu przykładu tekstu oraz oczekiwanej odpowiedzi modelu. W przeszłości jako cechy wykorzystywane były poszczególne słowa w tekście. Obecnie do takiej klasyfikacji tekstu używane są modele językowe, które uwzględniają dodatkowo zależności między wyrazami, co określamy

potocznie jako lepsze rozumienie tekstu. Są one również traktowane jako część rozwoju sztucznej inteligencji. W zakresie klasyfikacji treści tekstowych istotne są zatem dwa obszary, które zostaną opisane w tej publikacji: zbiory danych, które mogą służyć do treningu, oraz modele, które po wyuczeniu będą służyć do klasyfikacji tekstu.

Modele, które są obecnie stosowane do klasyfikacji treści, wywodzą się z dwóch modeli bazowych: BERT (*bidirectional encoder representations from transformers*) [Devlin i in., 2018] oraz GPT (*generative pre-trained transformer*) [Radford i in., 2018]. Oba wykorzystują architekturę transformatorów [Vaswani i in., 2017], która stała się punktem zwrotnym w zakresie rozwoju możliwości rozumienia tekstu. Stosują ją również duże modele językowe (*large language models*, LLM), utożsamiane często ze sztuczną inteligencją.

Zagadnienie klasyfikacji tekstu ze względu na potrzebę weryfikacji (*check-worthiness*) jest obszarem szczególnego zainteresowania jednej z konferencji poświęconej przetwarzaniu języka naturalnego – Conference and Labs of the Evaluation Forum [CLEFF, 2025]. Modele oparte na transformatorach, takie jak BERT oraz GPT i ich pochodne, były bardzo często wykorzystywane przez badaczy [Barrón-Cedeño i in., 2023; Nakov i in., 2022]. Polega to na tym, że podstawowy model jest dostrajany na odpowiednio dobranym zbiorze danych.

W 2022 r. najlepsze rozwiązanie zostało opracowane z wykorzystaniem modelu RoBERTa large [Liu i in., 2019], gdzie po dodatkowym procesie wzbogacania danych (dwukrotne tłumaczenie) uzyskano miarę F1 na poziomie 0,698 [Savchev, 2022]. F1 to średnia harmoniczna pomiędzy precyzją (*precision*) i czułością (*recall*), która jest często wykorzystywana do oceny modeli klasyfikacyjnych – są one tym lepsze, im wartość F1 jest bliższa 1,0. Kolejne rozwiązanie wykorzystywało złączenie 10 modeli (*ensemble*) bazujących na BERT oraz RoBERTa, osiągając F1 = 0,667 [Buliga Nicu, 2022]. Dopiero na trzecim miejscu z F1 = 0,626 był zespół, który wykorzystywał dostrojony model GPT-3 [Agrestia, Hashemianb, Carmanc, 2022]. Jest to istotne o tyle, że modele BERT ze względu na mniejszy rozmiar oraz otwartość mogą być uruchamiane lokalnie, natomiast GPT traktowane są jako wielkie modele i mogą być wykorzystywane jedynie w chmurze poprzez API, udostępnione przez OpenAI. Niemniej w następnym roku akurat rozwiązanie wykorzystujące GPT-3 curie, dotrenowane na ograniczonym, dobranym jakościowo zbiorze, okazało się najlepsze, osiągając F1 = 0,898 [Sawiński i in., 2023]. Ten sam zespół przedstawił również inne minimalnie słabsze rozwiązanie (F1 = 0,894), które z kolei wykorzystywało model lokalny DeBERTa v3 [He, Gao, Chen, 2021].

Istotnym zagadnieniem jest również przygotowanie zbioru danych. Stosowane są różne techniki, od wzbogacenia danych (*data augmentation*) celem zwiększenia

liczebności zbioru (np. tłumaczenie zbiorów z innych języków, dwukrotne tłumaczenie celem uzyskania tożsamej semantycznie i różnej leksykalnie treści), przez ekstrakcję dodatkowych cech (np. metadane z portalu-źródła informacji), uwzględnienie wektorowych osadzeń przygotowanych przy pomocy innych modeli, do dołączenia dodatkowych nieetykietowanych danych. Zadania w ramach CLEF nie ograniczają się do języka angielskiego, a dostępne zasoby w innych językach stwarzają dodatkową szansę na wzbogacenie zbiorów uczących, np. poprzez tłumaczenie maszynowe [Eyuboglu i in., 2022] lub wręcz wykorzystanie modeli wielojęzycznych [Mingzhe Du, Gollapalli, 2022]. Z innych badań wynika, że równie dobrze wyniki może poprawić celowane ograniczenie zbioru (tzw. *undersampling*) np. poprzez usunięcie przykładów niejednoznacznych i trudnych do nauczenia [Sawiński, Węcel, Księżniak, 2024b].

Co do zasady jednak, im więcej zasobów w danym języku i im bardziej są one różnorodne, tym większa szansa na wytrenowanie dobrego modelu. Różne badania wskazują, że mniej popularne języki nie znajdują się tutaj na straconej pozycji. Jak wskazuje Jiang i Zubiaga [2024], wiedza dziedzinowa z języków bogatszych w treść może być przeniesiona do tych mniej zasobnych, co określane jest jako Cross-Lingual Transfer Learning (CLTL). Szczególnie atrakcyjne jest wykorzystanie wielojęzycznych modeli językowych, dla których przykłady można podawać w dowolnym języku, niekoniecznie tym docelowym. Pomija się w ten sposób konieczność tłumaczenia tekstu – sam model wielojęzyczny zachowuje się jak translator. Możliwe jest „przeniesienie korzyści" nawet między tak odległymi językami jak arabski i niderlandzki [Sawiński, Węcel, Księżniak, 2024a]. Wspomniana publikacja jest o tyle interesująca, że testuje 15 wariantów zbiorów, utworzonych z wykorzystaniem różnych języków oraz czterech metod doboru próbek z populacji (*undersampling*).

7.2. Metodyka

Przeprowadzone wcześniej analizy literatury oraz bieżąca obserwacja dostępności zasobów do przetwarzania języka polskiego wskazują, że brakuje metod pozwalających na identyfikację stwierdzeń wymagających weryfikacji. Hipoteza wstępna jest taka, że luka ta wynika przede wszystkim z braku odpowiednio oznaczonych zasobów w języku polskim, na którym można by wytrenować modele. Nasze doświadczenia wskazują jednak, że dobre modele można dostroić nawet na zasobach w innych językach.

Celem naszych badań jest zatem opracowanie zbiorów danych oraz modeli językowych, które pozwoliłyby na identyfikację stwierdzeń w języku polskim, wymagających weryfikacji przez fact-checkerów. Z tego wynikają cele szczegółowe. Przede wszystkim

należy określić, jakie są dostępne korpusy tekstu z klasyfikacją według *check-worthy*. Jeśli nie ma wśród nich języka polskiego, należy taki zbiór opracować poprzez wybór odpowiednich stwierdzeń i ich ręczną anotację. Sposób anotacji zbioru powinien być określony w protokole anotacji. Wcześniej konieczne jest zdefiniowanie rozumienia *check-worthy*, tak aby każda z osób anotujących była świadoma, do jakich kryteriów należy się odwołać. W obszarze podcelów dotyczących modeli w pierwszej kolejności należy zidentyfikować modele dla języka polskiego oraz modele wielojęzyczne obsługujące nasz język. Kolejny krok to przygotowanie różnych wariantów zbiorów do uczenia. Po przeprowadzonych treningach możliwe będzie określenie, który wariant daje najdokładniejsze wyniki.

W toku prowadzonych prac zamierzamy odpowiedzieć na następujące pytania badawcze:

- P1. Jak opracować wysokiej jakości zbiór tekstów w języku polskim, służący do trenowania modeli klasyfikacji według konieczności weryfikacji?
- P2. W jaki sposób wykorzystać bogate zasoby w językach obcych do poprawy modelu dla języka polskiego?
- P3. Czy lepiej wykorzystywać dedykowane modele dla jednego języka (tutaj: polskiego), czy wystarczające są modele wielojęzyczne?
- P4. Jaka kombinacja zbioru danych oraz modelu daje najlepsze wyniki?
- P5. Czy racjonalne jest podejmowanie wysiłku w kierunku opracowania większego zbioru danych dla języka polskiego?

7.3. Opracowanie zbioru danych

W momencie prowadzenia badań nie był dostępny publicznie żaden zbiór tekstów w języku polskim z etykietami dotyczącymi konieczności weryfikacji (*check-worthiness*). Na potrzeby eksperymentów przygotowaliśmy własny zbiór, w skład którego weszły wiadomości w języku polskim, pochodzące z platformy społecznościowej X (dawniej Twitter), a także wykorzystaliśmy dwa istniejące angielskojęzyczne zbiory: ClaimBuster [Arslan i in., 2020] oraz zbiór testowy z konkursu CLEF2023 CheckThat! Lab Task 1b English [CheckThat, 2025]. Oba zbiory angielskojęzyczne zostały wykorzystane zarówno w wersji oryginalnej, jak i przetłumaczonej na język polski z wykorzystaniem translatora Google [Google Cloud, 2025].

7.4. Pozyskanie i wybór danych do procesu etykietowania

Teksty do polskiego zbioru danych zostały pozyskane z platformy X w 2023 r. przez dostępny wówczas interfejs API na licencji dla badań naukowych. Wybór odbył się na podstawie dwóch zestawów kryteriów, które w zamierzeniu miały wygenerować dwa odrębne zbiory, zawierające różny stosunek wiadomości wymagających sprawdzenia do wiadomości niewymagających sprawdzenia.

Pierwszy zbiór danych został pobrany przy użyciu zestawu kryteriów obejmujących listę 100 najpopularniejszych słów w języku polskim oraz ograniczenie dotyczące czasu publikacji wiadomości (lata 2019–2021), liczby interakcji z wiadomością (liczba wszystkich interakcji, tj. suma liczb odpowiedzi, przekazań, polubień i cytowań większa niż 100). Dodatkowo wybrane zostały tylko wiadomości, które nie zawierały dołączonych mediów audio lub wideo, nie cytowały, nie przekazywały ani nie odpowiadały na inną wiadomość. Intencją zastosowania takiego zestawu kryteriów było wybranie wiadomości tekstowych, które są kompletne (tzn. nie wymagają analizy kontekstu poprzednich wiadomości lub dołączonych plików medialnych) oraz posiadają znaczący potencjał wiralności, a jednocześnie stanowią reprezentatywną próbkę dla ogółu wiadomości z X w języku polskim. Ten zbiór danych będzie dalej nazywany „zbiorem bazowym".

Drugi zbiór danych został pobrany przy użyciu zestawu kryteriów, obejmujących ograniczenie dotyczące czasu publikacji wiadomości (lata 2022–2023), liczby interakcji z wiadomością (liczba wszystkich interakcji większa niż 100) oraz braku dołączonych mediów audio lub wideo. Dodatkowo wybrane zostały tylko wiadomości, które w odpowiedziach zawierały przynajmniej jedno słowo kluczowe: 'nieprawda', 'fałsz', 'fake', 'fake-news', 'fakenews', 'fejk'. Podobnie jak w przypadku zbioru bazowego wybrane zostały jedynie wiadomości, które nie cytowały, nie przekazywały ani nie odpowiadały na inną wiadomość, a także te, których autorzy w momencie ich pobierania mieli więcej niż 5000 obserwujących użytkowników. Intencją zastosowania takiego zestawu kryteriów było wybranie wiadomości tekstowych, które są kompletne oraz posiadają najwyższy potencjał wiralności (poprzez zastosowanie filtra popularności wiadomości oraz autora), a jednocześnie reakcje użytkowników wskazują na możliwość wystąpienia w niej dezinformacji. Ten zbiór danych będzie dalej nazywany „zbiorem filtrowanym reakcjami użytkowników".

Do etykietowania zostało wybranych 1201 przykładów ze „zbioru bazowego" oraz 1500 przykładów ze „zbioru filtrowanego reakcjami użytkowników".

7.5. Etykietowanie danych

Zadanie związane z etykietowaniem (anotacją) danych miało na celu zaklasyfikowanie własnego zbioru danych ze względu na zawieranie stwierdzeń, które powinny podlegać sprawdzeniu (*check-worthiness*). Zaanotowane dane miały zostać następnie wykorzystane w procesie trenowania i testowania modeli, opisanych w sekcji czwartej.

Proces anotacji składał się z kilku etapów, omówionych w kolejnych paragrafach:

1. Przegląd literatury związanej z procesem identyfikacji stwierdzeń wymagających sprawdzenia.
2. Przygotowanie przewodnika anotacji, uwzględniającego rodzaje klas wykorzystanych w procesie anotacji zbioru oraz zdefiniowanie procedur i zasad anotacji.
3. Przygotowanie środowiska do przeprowadzenia procesu anotacji.
4. Przeprowadzenie procesu anotacji, w tym uzgodnienie ostatecznych klasyfikacji w przypadku niezgodności między anotatorami.

Pierwszym krokiem procesu był przegląd literatury w zakresie definicji pojęcia „checkworthiness", a także kryteriów i dobrych praktyk stosowanych przez organizacje fact-checkingowe do identyfikacji stwierdzeń wymagających sprawdzenia. W tym celu zapoznano się z dokumentacją opisującą metodologie stosowane przez polskie i zagraniczne agencje fact-checkingowe, m.in. Stowarzyszenie Demagog [DEMAGOG, 2025], Fakehunter [FakeHunter, 2025], Politifact [POLITIFACT, 2025], AFP [AFP, 2025], International Fact-checking Network [IFCN Code of Principles, 2025]. Analiza dokumentacji wykazała, że pojęcie „check-worthy" nie jest jednoznacznie opisane w literaturze. Agencje fact-checkingowe często używają różnych kryteriów do określenia, czy dane stwierdzenie wymaga weryfikacji. Jednak punktem wspólnym jest tzw. „factual and verifiable claim", czyli takie stwierdzenie, które może zostać sprawdzone jako prawdziwe lub fałszywe na podstawie dostępnych dowodów, gdyż opiera się na rzeczywistych informacjach, danych, zdarzeniach i nie stanowi opinii, przekonania, prognozy czy interpretacji. Na przykład, według Demagoga, sprawdzeniu powinny podlegać treści merytoryczne (tj. zawierające odwołania do faktów, np. liczb, wydarzeń z przeszłości), które wychodzą poza zakres tzw. wiedzy powszechnej i są istotne z punktu widzenia debaty publicznej, mają znaczenie dla społeczeństwa. Fakehunter przy wyborze stwierdzeń do fact-checkingu bierze również pod uwagę to, czy stwierdzenie wydaje się mylące i nasuwa przypuszczenie niezgodności ze stanem faktycznym, a także czy jest szeroko udostępniane (ocena zasięgu internetowego oraz popularności źródła udostępniania informacji). W przypadku, gdy stwierdzenie dotyczy wypowiedzi znanych osób (np. w formie mowy zależnej), analizowane mogą być z kolei dwa aspekty: 1) czy wypowiedź osoby publicznej zawiera błędne lub niezgodne z rzeczywistością

stwierdzenia; 2) przypisanie danego oświadczenia osobie publicznej, które zostało sfabrykowane lub stanowi fałszywe przedstawienie jej wypowiedzi (stwierdzenia typu „Pan X powiedział, że…").

Przeprowadzona analiza dokumentacji umożliwiła identyfikację pewnych ogólnych zasad określających, czy dane stwierdzenie powinno (lub nie) podlegać sprawdzeniu (rozróżnienie między check-worthy i non check-worthy), a także zdefiniowanie dodatkowych klas do anotacji.

Klasy (etykiety), które zostały wykorzystane w procesie etykietowania zbioru, to:

- *Absent* – tekst nie zawiera stwierdzenia (*claim*) lub nie jest możliwa jego weryfikacja.
- *Not Check-Worthy* – tekst zawiera weryfikowalne stwierdzenie, ale nie jest ono warte
- weryfikacji.
- *General Public* – tekst zawiera stwierdzenie, które powinno zostać zweryfikowane, ponieważ jest istotne z punktu widzenia debaty publicznej. Do tej kategorii zaliczane są stwierdzenia związane z polityką (np. wypowiedzi polityków, wyniki sondaży), technologią, sportem, wydarzeniami społecznymi.
- *Harmful* – tekst zawiera stwierdzenie, które powinno zostać zweryfikowane, ponieważ może bezpośrednio narazić czyjeś zdrowie lub życie. Do tej kategorii nie są jednak zaliczane stwierdzenia, które nie wpisują się w ogólną narrację (tzn. nie dotyczą nośnych tematów, np. COVID-19).
- *General Public and Harmful* – tekst zawiera stwierdzenie, które powinno zostać zweryfikowane, ponieważ może narazić czyjeś zdrowie i dotyczy szerokiego grona odbiorców (np. stwierdzenie związane z nośnymi tematami, takimi jak COVID-19, czy dotyczące ataków na mniejszości narodowe).

Wspomnianą wyżej narrację ogólną należy rozumieć jako zbiór powiązanych semantycznie stwierdzeń, pojawiających się w wiadomościach publikowanych przez serwisy internetowe ze względnie dużą częstością, o dużej wiralności, tzn. szybko rozprzestrzeniające się. Przykładowo, po wpisaniu do wyszukiwarki Google stwierdzenia: „szczepienia na covid wywołują autyzm" można bez problemu znaleźć wiadomości, w których rozważana jest jego prawdziwość – wokół tego tematu powstało wiele artykułów i opublikowano wiele postów w mediach społecznościowych, w których podjęto w związku z nim dyskusje. Kontrprzykładem jest natomiast stwierdzenie: „jedzenie proszku do prania pomaga na trawienie" – próba wyszukania powiązanych z nim artykułów czy postów w mediach społecznościowych nie przynosi rezultatów, niemniej stwierdzenie bezsprzecznie nawołuje do działań zagrażających życiu i zdrowiu.

Do najważniejszych wytycznych w kwestii identyfikacji stwierdzeń, które nie powinny podlegać weryfikacji, zaliczyliśmy te: 1) będące opiniami, 2) związane z osobistym

doświadczeniem, 3) odnoszące się do powszechnej wiedzy, 4) zawierające predykcje na przyszłość, 5) związane z nieistniejącymi rzeczami/zjawiskami, 6) odnoszące się do wartości estetycznych, moralnych lub etycznych, 7) dotyczące wiary lub zdolności nadprzyrodzonych, 8) wyrażające silne emocje, reakcje lub stany bez formalnego deklarowania faktów, intencji czy zobowiązań, 9) będące tautologiami w sensie językowym, np. kawalerowie nie mają żon. Informacje te zostały następnie zebrane w formie przewodnika dla anotatorów.

W procesie etykietowania brały udział trzy osoby – członkowie zespołu projektowego OpenFact, mające kilkuletnie doświadczenie badawcze w zakresie analizy tekstów zawierających fałszywe informacje. Dwie z nich pełniły rolę anotatorów, których zadaniem była niezależna klasyfikacja każdego rekordu w zbiorze, natomiast trzecia osoba pełniła rolę recenzenta oceniającego te przypadki, dla których nie było zgodności między dwoma anotatorami. W przypadku tekstów granicznych (co do których recenzent miał wątpliwości dotyczące właściwej klasyfikacji) odbywała się dodatkowa dyskusja w zespole trzyosobowym. Przykładami takich tekstów są: „W 2021 r., jako nauczyciele zarabialiśmy tak mało, że nawet uczniowie się z nas śmiali. Jeden nauczyciel wykonuje pracę za trzech (jeżeli tylko ma papier), bo tyle chętnych do pracy w szkole…" oraz „Im więcej rozmawiam i czytam na ten temat, tym bardziej utwierdzam się w przekonaniu, że za drakońskie podwyżki cen gazu odpowiada PiS i PGNiG, który stracił miliardy zł z powodu złych decyzji dotyczących formuły cenowej".

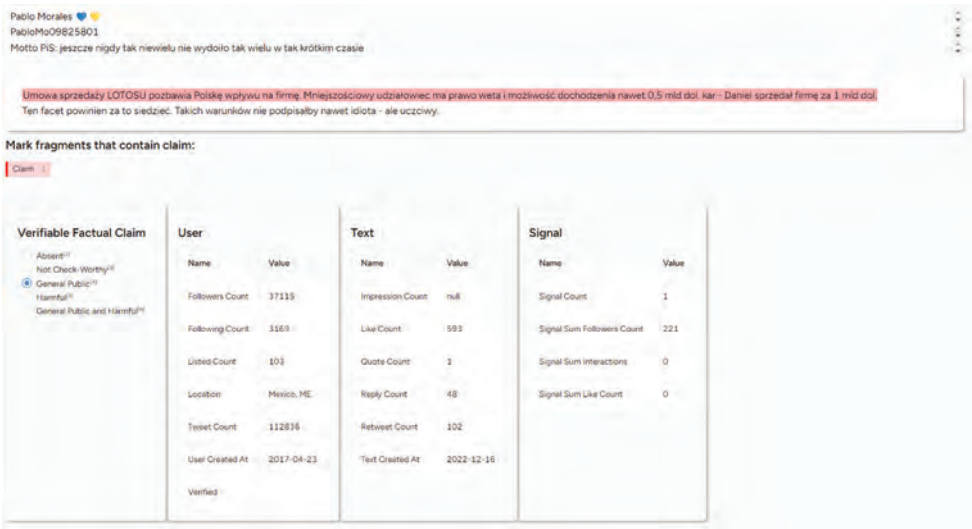
Zakres danych udostępnionych anotatorom, oprócz samej treści posta, który miał zostać zaklasyfikowany, obejmował również informacje o użytkowniku, który daną treść opublikował (nazwa konta, liczba osób śledzących, liczba profili śledzonych przez dane konto, lokalizacja, liczba opublikowanych tweetów, data utworzenia profilu, a także informacja, czy konto zostało zweryfikowane przez platformę X), statystyki dotyczące posta (liczba wyświetleń, polubień, cytowań, odpowiedzi, retweetów, data stworzenia posta), statystyki dotyczące tzw. sygnałów[1] (liczba sygnałów, suma liczb śledzących dla kont, od których sygnał pochodzi, liczba wyświetleń i polubień sygnału).

Do procesu anotacji wykorzystano oprogramowanie LabelStudio [Label Studio, 2025] w wersji Enterprise, które oprócz udostępnienia anotatorom opisanych wyżej danych oraz możliwości przeprowadzenia procesu etykietowania (rysunek 7.1) umożliwiło również zarządzanie całym procesem anotacji (publikacja przewodnika dla anotatorów, przypisanie anotatorów do zadań, analiza wyników anotacji

[1] Sygnał to odpowiedź na oryginalny post, np. w formie komentarza, sugerująca, że dana treść jest nieprawdziwa. Sygnały są identyfikowane na podstawie zdefiniowanej listy słów kluczowych, takich jak: 'nieprawda', 'kłamstwo', 'fake news'.

i identyfikacja braku zgodności, dodawanie komentarzy do przykładów wątpliwych). W przypadku dłuższych tekstów narzędzie pozwalało również na zaznaczenie jego fragmentu, stanowiącego ocenione stwierdzenie (czerwony fragment tekstu widoczny na rysunku 7.1).

Rysunek 7.1. Narzędzie LabelStudio wykorzystane w procesie anotacji danych



Źródło: opracowanie własne.

7.6. Przygotowanie zbiorów do treningu modeli językowych

Dane pochodzące ze zbiorów ClaimBuster, CLEF2023 CheckThat! Lab oraz wiadomości z X zostały podzielone na podzbiory, których następnie użyto do skonstruowania zbiorów treningowych, ewaluacyjnych i testowych dla treningu modeli językowych.

Podstawę treningu stanowiły dane ze zbioru ClaimBuster, który zawiera 23 533 przykłady z etykietami, przy czym liczebność stwierdzeń z etykietą *check-worthy* (w dalszej części określanej jako klasa pozytywna) wynosi 5485. Poprzednie badania [Sawiński, Węcel, Księżniak, 2024b] pokazały, że redukcja liczebności próby klasy większościowej zapewnia efektywny trening modelu. Dysponując informacją o jakości etykiet, do zbioru treningowego włączono wszystkie przykłady pozytywne oraz taką samą liczbę przykładów negatywnych, wylosowanych spośród tekstów o etykietach wysokiej jakości. Zbiór będzie dalej nazywany *ClaimBuster-en*, a jego przetłumaczona na polski wersja będzie dalej nazywana *ClaimBuster-pl*.

Zbiór testowy z konkursu CLEF2023 CheckThat! Lab Task 1b English, zawierający 315 przykładów, został dołączony jako dodatkowy zbiór testowy, aby móc odnieść wyniki eksperymentów do wyników konkursu. Oryginalne teksty w języku angielskim zostały użyte do stworzenia zbioru nazywanego dalej *CheckThat23-en*. Przetłumaczone teksty posłużyły do stworzenia analogicznego zbioru w języku polskim, nazywanego dalej *CheckThat23-pl*. Wszystkie 1201 przykładów z ręcznie przygotowanymi etykietami ze zbioru bazowego zostały włączone do podzbioru treningowego, nazywanego dalej BAZA. Przykłady ze zbioru filtrowanego reakcjami użytkowników podzielono losowo na 3 równe grupy o liczebności 500, z których pierwsza została podzbiorem zbioru treningowego, nazywanym dalej FILTR. Druga grupa przykładów została zbiorem ewaluacyjnym, nazywanym dalej EWALUACJA, a trzecia – zbiorem testowym, nazywanym dalej TEST.

Złączenia różnych podzbiorów posłużyły do skonstruowania ośmiu wariantów zbioru treningowego. Cztery z nich składały się wyłącznie z przykładów w języku polskim i były użyte do treningu wszystkich modeli: 1) ClaimBuster-pl, 2) ClaimBuster-pl + BAZA, 3) ClaimBuster-pl + FILTR, 4) ClaimBuster-pl + BAZA + FILTR.

Pozostałe warianty składały się w głównej mierze z przykładów w języku angielskim, do których były dokładane mniejsze zbiory w języku polskim. Warianty te zostały użyte wyłącznie do treningu modeli wielojęzycznych: 1) ClaimBuster-en, 2) ClaimBuster-en+BAZA, 3) ClaimBuster-en + FILTR, 4) ClaimBuster-en + BAZA + FILTR.

W każdym treningu do ewaluacji użyto tego samego zbioru, oznaczonego jako *EWALUACJA*. W przypadku treningu wszystkich modeli do testów wykorzystano dwa zbiory polskojęzyczne, oznaczone jako *TEST* i *CheckThat2023-pl*. Dodatkowo dla modeli wielojęzycznych miary jakości klasyfikacji były obliczane dla angielskojęzycznego zbioru *CheckThat2023-en*.

Szczegóły zostały zaprezentowane w tabeli 7.1.

Tabela 7.1. Zbiory danych treningowych, ewaluacyjnych i testowych

Nazwa podzbioru	Język	Liczba przykładów	Stosunek przykładów pozytywnych do negatywnych	Źródło	Anotacja
BAZA	PL	1201	1:3.5	X, *zbiór bazowy*	własna
FILTR	PL	500	1:1.7	X, *zbiór filtrowany reakcjami użytkowników*	własna
EWALUACJA	PL	500	1:1.6		
TEST	PL	500	1:1.8		
ClaimBuster-en	EN	10 970	1:1	debaty prezydenckie w USA	zbiorowa
ClaimBuster-pl	PL				

Nazwa podzbioru	Język	Liczba przykładów	Stosunek przykładów pozytywnych do negatywnych	Źródło	Anotacja
CheckThat23-en	EN	315	1:1.9	brak danych	brak danych
CheckThat23-pl	PL				

Źródło: opracowanie własne.

7.7. Opracowanie modeli

7.7.1. Wybór modeli bazowych

Wyboru modeli wykorzystanych do treningu dokonano na podstawie trzech kryteriów:

- **Wyniki osiągnięte w benchmarku KLEJ** – benchmark KLEJ (Kompleksowa Lista Ewaluacji Językowych) jest zestawem testów oceniających zdolności modeli przetwarzania języka naturalnego w zakresie zadań obejmujących m.in. klasyfikację tekstu czy analizę wydźwięku [Rybak i in., 2020] (Kryterium 1: K1).
- **Zasoby sprzętowe** – eksperymenty przeprowadzano na serwerze wyposażonym w cztery karty NVIDIA GeForce RTX 2080 Ti z pamięcią 11 GB na kartę; zasoby te nie pozwoliły na uruchomienie niektórych większych modeli językowych (Kryterium 2: K2).
- **Specyfika modeli** – ze względu na ogólny charakter zadania klasyfikacji (szeroki zakres tematyczny twierdzeń, takich jak zdrowie, polityka, nastroje społeczne) nie wybierano modeli przystosowanych do specyficznych zadań (np. dotrenowanych do detekcji cyberagresji) (Kryterium 3: K3).

Tabela 7.2 przedstawia wyniki osiągnięte przez różne modele według poszczególnych kryteriów. Na tej podstawie zdecydowano się na przeprowadzenie treningu z wykorzystaniem modelu wielojęzycznego XLM-RoBERTa base [Conneau i in., 2019] i dwóch modeli jednojęzycznych: HerBERT base [Mroczkowski i in., 2021], Polish RoBERTa v2 base [Dadas, Perełkiewicz, Poświata, 2020]. Dodatkowo podjęto decyzję o przeprowadzeniu treningu z wykorzystaniem – nieujętego w benchmarku KLEJ – wielojęzycznego modelu mDeBERTa v3 base [He i in., 2020], który dzięki zastosowanej architekturze poprawia wyniki testu w zadaniu XNLI dla 15 języków, osiągając średni wynik 79,8 w porównaniu z 76,2 osiągniętym przez wielojęzyczny model RoBERTa [Dadas, Perełkiewicz, Poświata, 2020].

Tabela 7.2. Wyniki benchmarku KLEJ dla różnych modeli

Nazwa modelu	Wynik benchmarku KLEJ (K1)	Zasoby sprzętowe (K2)	Charakterystyka zastosowań (K3)
Polish RoBERTa v2 large	88,9	nie spełnia	spełnia
HerBERT large	88,4	nie spełnia	spełnia
XLM-RoBERTa large + NKJP	87,8	nie spełnia	spełnia
Polish RoBERTa large	87,8	nie spełnia	spełnia
XLM-RoBERTa large	87,5	nie spełnia	spełnia
Polish RoBERTa-v2 base	**86,8**	**spełnia**	**spełnia**
XLM-RoBERTa large	86,6	nie spełnia	spełnia
HerBERT base	**86,3**	**spełnia**	**spełnia**
TrelBERT	86,1	spełnia	nie spełnia
Polish RoBERTa base	85,3	spełnia	spełnia
XLM-RoBERTa large	84,7	nie spełnia	spełnia
PoLitBert v32k linear 50k (kilka rodzajów)	83,5	spełnia	nie spełnia
Sup-SimCSE SNLI XLM-R base	81,7	spełnia	nie spełnia
Polbert (cased)	81,7	spełnia	spełnia
XLM-RoBERTa base	**81,5**	**spełnia**	**spełnia**

Źródło: opracowanie własne.

7.7.2. Trenowanie modeli

Wskazane wyżej modele zostały dostrojone na podstawie opracowanych zbiorów danych z uwzględnieniem następujących wartości hiperparametrów: learning rate (testowano wartości: 2e-6, 3e-6, 5e-6, 1e-5) oraz optimizer (adamw torch dla wszystkich wybranych modeli i dodatkowo rmsprop oraz nadam dla modeli: mDeBERTa v3 base, HerBERT base).

Wykonano optymalizację hiperparametrów metodą bayesowską z wykorzystaniem funkcji sweep, dostępnej w narzędziu Weights & Biases [2025]. Metoda ta polega na tworzeniu probabilistycznego modelu funkcji celu, którą chcemy zminimalizować lub zmaksymalizować (przyjęto konfigurację: maksymalizacja wartości funkcji F1 dla klasy pozytywnej) podczas trenowania modelu. Proces ten polega na iteracyjnym wyborze kombinacji hiperparametrów, trenowaniu modelu, ocenie wyników i aktualizacji przekonań na temat najlepszych wartości hiperparametrów. Łącznie wykonano 192 próby w poszukiwaniu optymalnych wartości hiperparametrów. Najlepsze wyni-

ki osiągnięto z zastosowaniem optymalizatora „adamw torch" oraz wartości learning rate: 1e-5 (model: herbert base oraz Polish RoBERTa base), 2e-6 (model XLM-RoBERTa base), 5e-6 (model mDeBERTa v3 base). Wskazane hiperparametry zostały użyte do przeprowadzenia ostatecznych eksperymentów.

7.8. Doświadczenia z przeprowadzonych eksperymentów

Zaprezentowane wyniki zostały zebrane w trakcie 240 treningów modeli językowych, przy 24 wariantach kombinacji rodzaju modelu i zbioru treningowego oraz 10 wariantach losowej inicjalizacji modelu.

We wszystkich przypadkach treningu modeli jednojęzycznych model HerBERT base osiągnął wyższe wyniki F1 dla pozytywnej klasy niż model Polish RoBERTa v2 base podczas porównania zarówno wartości maksymalnych, jak i średnich osiągniętych w kolejnych treningach z różną losową inicjalizacją. Podobnie we wszystkich przypadkach treningów modeli wielojęzycznych model mDeBERTa v3 base osiągnął wyższe wyniki F1 dla pozytywnej klasy niż model XLMRoBERTa base podczas porównania zarówno wartości maksymalnych, jak i średnich osiągniętych w kolejnych treningach z różną losową inicjalizacją (tabela 7.3).

Tabela 7.3. Maksymalne i średnie wartości F1 dla klasy pozytywnej, wyliczone dla zbioru testowego TEST według wybranych czterech modeli trenowanych z dziesięcioma różnymi losowymi inicjalizacjami parametrów (CB = ClaimBuster)

Model	HerBERT base		Polish RoBETa v2 base		mDeBERTa v3 base		XLM-RoBERTa base	
Zbiór treningowy	maks.	śr.	maks	śr.	maks.	śr.	maks.	śr.
CB-pl	72,73	70,80	72,59	69,91	72,35	70,77	69,87	67,21
CB-pl + BAZA	75,42	73,42	73,76	71,86	74,44	72,13	72,11	70,80
CB-pl + FILTR	76,68	75,13	74,36	72,15	75,53	72,89	73,49	71,81
CB-pl + BAZA + FILTR	79,56	76,25	76,57	74,29	75,74	74,70	74,81	72,39
CB-en	–	–	–	–	73,26	69,32	71,03	68,84
CB-en + BAZA	–	–	–	–	77,11	73,91	72,90	71,50
CB-en + FILTR	–	–	–	–	77,30	75,96	75,44	74,21
CB-en + BAZA + FILTR	–	–	–	–	79,78	77,30	77,60	75,34

Źródło: opracowanie własne.

Zastosowanie do treningu wyłącznie danych ze zbioru ClaimBuster pozwalało osiągnąć na polskim testowym zbiorze *TEST* najlepszy wynik F1 na poziomie 73,26 dla zbioru treningowego w języku angielskim oraz 72,73 dla zbioru w języku polskim.

Wyniki osiągnięte na angielskojęzycznym zbiorze testowym *CheckThat23-en* wyniosły 91,08, czyli o 2% więcej niż zwycięski model zgłoszony na ten konkurs, gdy do treningu została użyta angielska wersja zbioru ClaimBuster, a tylko 88,12, gdy model był trenowany na przetłumaczonej na polski wersji zbioru. Wyniki osiągnięte na tym samym zbiorze testowym przetłumaczonym na polski *CheckThat23-pl* wyniosły 88,12, gdy do treningu została użyta angielska wersja zbioru ClaimBuster i aż 91,35, gdy model był trenowany na wersji zbioru przetłumaczonej na polski. Podobna, około trzyprocentowa, różnica w obu przypadkach przemawiała na korzyść stosowania zbioru treningowego i testowego w tym samym języku.

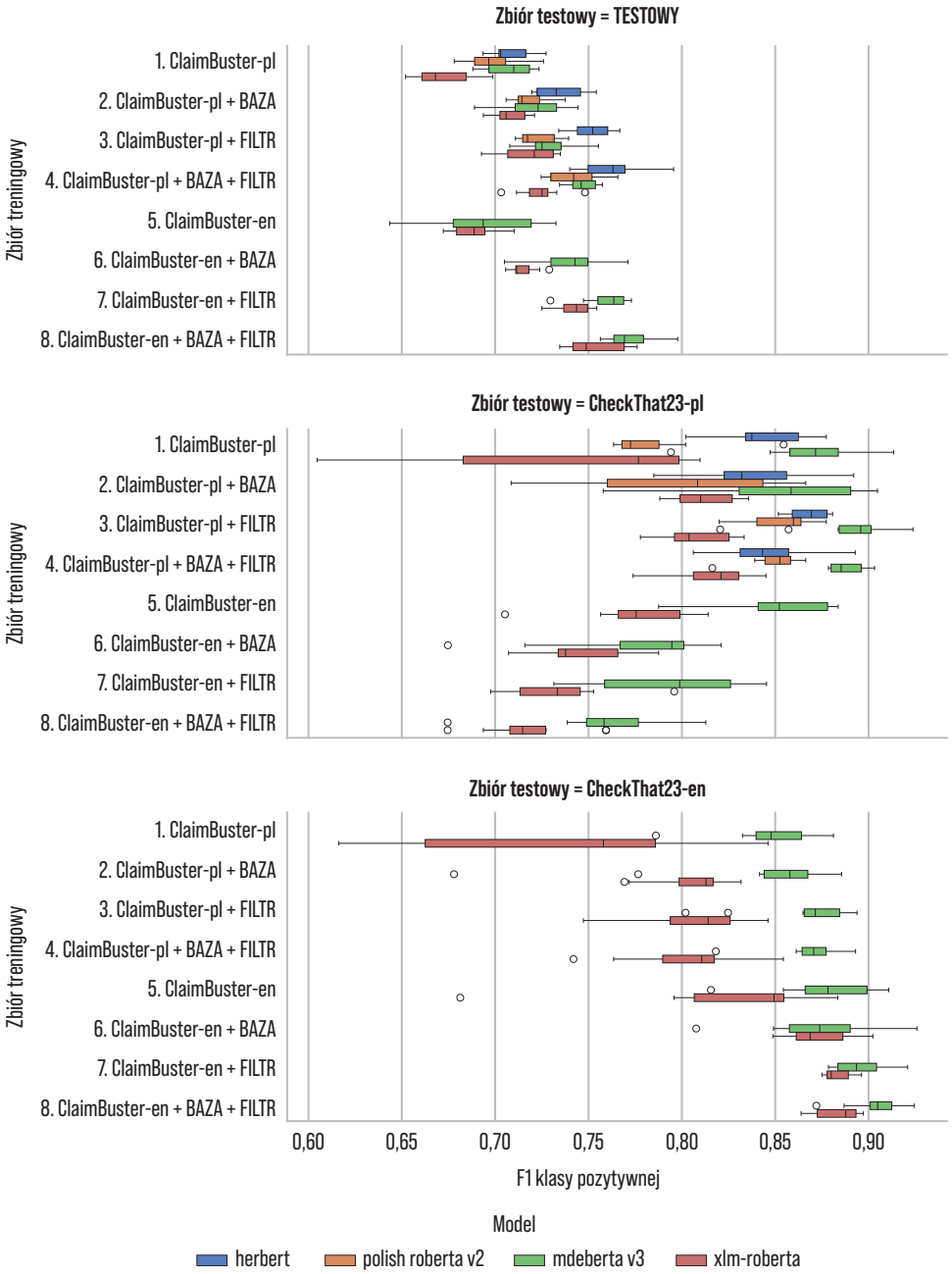
Rozszerzenie zbioru treningowego o dodatkowe przykłady w języku polskim poprawiało wynik na polskim testowym zbiorze TEST. Przy treningu na angielskojęzycznym ClaimBusteren dodanie zbioru BAZA poprawiało średnio wynik o 3,62%, dodanie zbioru FILTR poprawiało średnio wynik o 6,0%, a dodanie obu zbiorów BAZA i FILTR poprawiało średnio wynik o 7,24%. Podobne wyniki na zbiorze testowym CheckThat23-en zostały osiągnięte przy treningu na obu zbiorach ClaimBuster-pl oraz ClaimBuster-en. Jednocześnie rozszerzenie zbioru treningowego o dodatkowe przykłady w języku polskim pogarszało wynik na polskim testowym zbiorze CheckThat23-pl przy treningu na angielskojęzycznym ClaimBuster-en. Spadki wynosiły średnio 5,24% po dodaniu, dodanie zbioru FILTR pogarszało średnio wynik o 4,96%, a dodanie obu zbiorów BAZA i FILTR pogarszało średnio wynik o 7,49%. Rozszerzenie zbioru treningowego ClaimBuster-en poprawiało wyniki dla CheckThat23-pl jak w poprzednich przypadkach. Uśrednione wyniki przedstawione zostały w tabeli 7.4, a wyniki w formie wizualnej na rysunku 7.2.

Tabela 7.4. Maksymalne i średnie wartości F1 dla klasy pozytywnej, wyliczone dla trzech zbiorów testowych według wybranych czterech modeli trenowanych z dziesięcioma różnymi losowymi inicjalizacjami parametrów (CB = ClaimBuster)

Zbiór testowy	Średni wynik F1			Maksymalny wynik F1		
Zbiór treningowy	CheckThat 23-en	CheckThat 23-pl	TEST	CheckThat 23-en	CheckThat 23-pl	TEST
CB-pl	78,96	81,10	69,67	88,12	91,35	72,73
CB-pl + BAZA	82,06	82,83	72,05	88,56	90,48	75,42
CB-pl + FILTR	83,85	85,39	73,00	89,38	92,38	76,68
CB-pl + BAZA + FILTR	83,62	84,87	74,41	89,30	90,32	79,56
CB-en	85,19	**81,28**	69,08	91,08	88,37	73,26
CB-en + BAZA	87,32	**76,04**	72,70	92,59	82,11	77,11
CB-en + FILTR	88,94	**76,32**	75,08	92,09	84,54	77,30
CB-en + BAZA + FILTR	89,32	**73,79**	76,32	92,45	81,28	79,78

Źródło: opracowanie własne.

Rysunek 7.2. **Wykresy pudełkowe miar F1 dla klasy pozytywnej, wyliczonych dla trzech różnych zbiorów testowych (poszczególne wykresy) według czterech różnych modeli (kolor) trenowanych na ośmiu różnych zbiorach treningowych (wiersze w grupie)**



Źródło: opracowanie własne.

Najlepsze wyniki F1 na polskim testowym zbiorze *TEST*, czyli 79,78%, osiągnął model mDeBERTa v3 base z wykorzystaniem do treningu połączonych zbiorów *ClaimBuster-en* oraz *BAZA* i *FILTR*. Tylko nieznacznie gorszy wynik, 79,56%, osiągnął model HerBERT base trenowany z użyciem *ClaimBuster-pl* oraz *BAZA* i *FILTR*.

Podsumowanie

Przeprowadzona analiza literatury wskazała na niejednoznaczne rozumienie pojęcia „stwierdzenie wymagające weryfikacji" (*check-worthy*). Zebrane wskazówki oraz wcześniejsze doświadczenia pozwoliły jednak na opracowanie spójnego przewodnika anotacji i utworzenie własnego zbioru tekstów wysokiej jakości, co odpowiada na pierwsze pytanie badawcze (P1).

Istniejące zbiory w językach obcych mogą być wykorzystane na dwa sposoby, których skuteczność została potwierdzona w toku prowadzonych badań. Pierwszy z nich to maszynowe tłumaczenie zbioru i wykorzystanie do trenowania dedykowanego modelu w języku polskim – potwierdzone dla zbioru ClaimBuster-pl i modelu HerBERT. Drugi zaś to bezpośrednie wykorzystanie dla trenowania modelu wielojęzycznego, co potwierdziliśmy dla ClaimBuster-en i modelu mDeBERTa v3 base. Wspomniane sposoby odpowiadają na pytanie badawcze P2. Przeprowadzone badania wskazują, że oba modele osiągnęły porównywalną skuteczność, mierzoną wynikiem F1 dla klasy pozytywnej i wynoszącą w zaokrągleniu 80%. Odpowiadając na trzecie pytanie badawcze P3, należy wskazać, że korzystniejsze jest użycie modeli wielojęzycznych: unikamy konieczności tłumaczenia zasobów na język polski oraz pozostawiamy sobie możliwość wykorzystania większych i bardziej różnorodnych zasobów w językach innych niż polski.

Aby uzyskać odpowiedź na czwarte pytanie badawcze P4, analizując kombinacje zbioru danych oraz modelu pod kątem najlepszych rezultatów uczenia maszynowego, należy w pierwszej kolejności odnieść się do składu zbiorów. Niezależnie od tego, czy były trenowane modele jedno- czy wielojęzyczne, najlepsze wyniki osiągnęliśmy podczas wykorzystania kombinacji trzech zbiorów: ClaimBuster oraz własnych etykietowanych ręcznie danych z X, zarówno tych wyszukanych na podstawie popularnych słów kluczowych, jak i filtrowanych reakcjami użytkowników. Ciekawa jest również obserwacja, że modele wielojęzyczne uczyły się lepiej, jeśli były trenowane na danych w dwóch językach, w porównaniu z zasobami jednojęzycznymi. Ponownie należy zauważyć, że różnorodność dała przewagę. Jeśli weźmiemy pod uwagę same mode-

le wielojęzyczne, okazuje się, że praktycznie we wszystkich eksperymentach model mDeBERTa v3 base osiągnął lepsze rezultaty niż model XLM-RoBERTa base.

W toku prowadzonych badań stwierdziliśmy, że zwiększanie liczebności i różnorodności zbioru jest zasadne. Z danych zgromadzonych w tabeli 7.3 wynika, że zwiększanie liczebności zbioru treningowego prowadzi do zwiększenia miary F1 na różnych zbiorach testowych, praktycznie niezależnie od użytego modelu. Należy również zauważyć, że modele trenowane ze zbiorem FILTR, tj. wybrane z uwzględnieniem reakcji użytkowników, osiągają lepsze wyniki niż BAZA, wybrane losowo. A zatem, odpowiadając na pytanie badawcze P5, należy stwierdzić, że warto opracować większy zbiór dla języka polskiego z bardzo dobrymi jakościowo anotacjami. Modele wielojęzyczne mają tę dodatkową przewagę, że mogą korzystać z większych zasobów, bez konieczności tłumaczenia. Wynika z tego, że powinno się poszukiwać podobnych zasobów w innych językach, gdyż również one przyczyniają się do lepszych wyników modeli wielojęzycznych do stosowania dla języka polskiego. Rozbudowa zbioru dla języka polskiego nie musi ograniczać się do zasobów wyłącznie polskich.

Wnioski z przeprowadzonych badań ograniczają się do języka polskiego oraz angielskiego. Z innych naszych prac wynika jednak, że wzbogacać mogą się tak odległe języki jak niderlandzki i arabski. Zauważamy również pewne różnice w zbiorach testowych używanych w ramach ClaimBuster oraz opracowanych przez nas (TEST) – ten drugi wydaje się większym wyzwaniem dla modeli. Kolejnym ograniczeniem jest liczebność zbiorów wysokiej jakości, gdzie ręcznie ocenione zostało 1500 dokumentów. W najbliższym czasie liczbę tę zwiększymy docelowo do 15 000, z wykorzystaniem metodyki opracowanej w ramach niniejszych badań. Modele będą również trenowane z większą liczbą losowych inicjalizacji parametrów (tzw. *seed*), gdyż zauważyliśmy, że wartość F1 może znacznie się wahać ze względu na wpadanie w różne minima lokalne.

Bibliografia

AFP (2025). *AFP Fact-Checking Stylebook*, https://sprawdzam.afp.com/node/54477 (dostęp: 30.03.2025.

Agrestia, S., Hashemianb, A., Carmanc, M. (2022). *PoliMi-FlatEarthers at CheckThat! 2022: GPT-3 Applied to Claim Detection*, "Working Notes of CLEF 2022" – Conference and Labs of the Evaluation Forum.

Arslan, F., Hassan, N., Li, C., Tremayne, M. (2020). *ClaimBuster: A Benchmark Dataset of Check-worthy Factual Claims*. DOI: 10.5281/zenodo. 3836810.

Barrón-Cedeño, A., Alam, F., Galassi, A., Da San Martino, G., Nakov, P., Elsayed, T., Azizov, D., Caselli, T., Cheema, G.S., Haouari, F., Hasanain, M., Kutlu, M., Li, C., Ruggeri, F., Struß, J.M., Zaghouani, W. (2023). Overview of the CLEF – 2023 CheckThat! Lab on Checkworthiness, Subjectivity, Political Bias, Factuality, and Authority of News Articles and Their Source. W: *Experimental IR Meets Multilinguality, Multimodality, and Interaction* (s. 251–275), A. Arampatzis, E. Kanoulas, T. Tsikrika, S. Vrochidis, A. Giachanou, D. Li, M. Aliannejadi, M. Vlachos, G. Faggioli, N. Ferro (Eds.). Cham: Springer. DOI: 10.1007/978-3-031-42448-9.

Buliga Nicu, R. (2022). *Zorros at CheckThat! 2022: Ensemble Model for Identifying Relevant Claims in Tweets*, "Working Notes of CLEF 2022" – Conference and Labs of the Evaluation Forum.

Cazalens, S., Leblay, J., Lamarre, P., Manolescu, I., Tannier, X. (2018). Computational fact checking: a content management perspective, *Proceedings of the VLDB Endowment*, 11(12), s. 2110–2113.

CheckThat (2025). *CT23_1B_checkworthy_english_test.zip*, https://gitlab.com/checkthat_lab/clef2023-checkthat-lab/-/blob/main/task1/data/CT23_1B_checkworthy_english_test.zip (dostęp: 30.03.2025).

CLEF (2025). *Conference and Labs of the Evaluation Forum*, http://www.clef-initiative.eu/ (dostęp: 30.03.2025).

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., Stoyanov, V. (2019). *Unsupervised Cross-lingual Representation Learning at Scale*. DOI: 10.48550/arXiv.1911.02116.

Dadas, S., Perełkiewicz, M., Poświata, R. (2020). Pre-Training Polish Transformer-Based Language Models at Scale. W: *Artificial Intelligence and Soft Computing* (s. 301–314), L. Rutkowski, R. Scherer, M. Korytkowski, W. Pedrycz, R. Tadeusiewicz, J.M. Zurada (Eds.). Cham: Springer. DOI: 0.1007/978-3-030-61534-5_27.

DEMAGOG (2025). *Metodologia*, https://demagog.org.pl/metodologia/ (dostęp: 30.03.2025).

Devlin, J., Chang, M., Lee, K., Toutanova, K. (2018). *BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding*. DOI: 10.48550/arXiv.1810.04805.

Eyuboglu, A., Arslan, M., Sonmezer, E., Kutlu, M. (2022). *TOBB ETU at CheckThat! 2022: Detecting Attention-Worthy and Harmful Tweets and Check-Worthy Claims*, "Working Notes of CLEF 2022" – Conference and Labs of the Evaluation Forum.

FakeHunter (2025). *Reguły weryfikacji faktów*, https://fake-hunter.pap.pl/sites/default/files/2022-07/reguly_weryfikacji_faktow.pdf (dostęp: 30.03.2025).

Google Cloud (2025). *Translate Documents, Websites, Apps, Audio Files, Videos with Support for More than 135 Languages*, https://cloud.google.com/translate (dostęp: 30.03.2025).

He, P., Gao, J., Chen, W. (2021). *DeBERTaV3: Improving DeBERTa using ELECTRA-Style Pre-Training with Gradient-Disentangled Embedding Sharing*. DOI: 10.48550/arXiv.2111.09543.

He, P., Liu, X., Gao, J., Chen, W. (2020). *DeBERTa: Decoding-Enhanced BERT with Disentangled Attention*. DOI: 10.48550/arXiv.2006.03654.

IFCN Code of Principles (2025). *The Commitments of the Code of Principles*, https://ifcncodeofprinciples.poynter.org/the-commitments (dostęp: 29.04.2025).

Jiang, A., Zubiaga, A. (2024). *Cross-Lingual Offensive Language Detection: A Systematic Review of Datasets, Transfer Approaches and Challenges*. DOI: 10.48550/arXiv.2401.09244.

Label Studio (2025). *Open Source Data Labeling Platform*, https://labelstud.io/ (dostęp: 30.03.2025).

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V. (2019). *RoBERTa: A Robustly Optimized BERT Pretraining Approach*. DOI: 10.48550/arXiv.1907.11692.

Micallef, N., Armacost, V., Memon, N., Patil, S. (2022). True or False: Studying the Work Practices of Professional Fact-Checkers, *Proceedings of the ACM on Human-Computer Interaction*, 6 (CSCW1), s. 1–44.

Mingzhe Du, S., Gollapalli, S.D. (2022). *NUS-IDS at CheckThat! 2022: Identifying Checkworthiness of Tweets Using ChecktaT5*, "Working Notes of CLEF 2022" – Conference and Labs of the Evaluation Forum.

Mroczkowski, R., Rybak, P., Wróblewska, A., Gawlik, I. (2021). HerBERT: Efficiently Pretrained Transformer-based Language Model for Polish. W: *Proceedings of the 8th WorkshoponBalto-SlavicNaturalLanguageProcessing* (s. 1–10). Kiyv: Association for Computational Linguistics, https://www.aclweb. org/anthology/2021.bsnlp-1.1 (dostęp: 30.03.2025).

Nakov, P., Barrón-Cedeño, A., Da San Martino, G., Alam, F., Kutlu, M., Zaghouani, W., Li, C., Shaar, S., Mubarak, H., Nikolov, A. (2022). Overview of the CLEF-2022 CheckThat! Lab Task 1 on Identifying Relevant Claims in Tweets. W: *CLEF 2022. Conference and Labs of the Evaluation Forum* (s. 368–392). Bologna: CEUR Workshop Proceedings.

POLITIFACT (2025). *Here's a Fact: You Ended up in the Wrong Place!*, https://www.politifact.com/article/2018/feb/12/principles-truth-o-meter-politifacts-methodology- (dostęp: 30.03.2025).

Radford, A., Narasimhan, K., Salimans, T., Sutskever, I. (2018). *Improving Language Understanding by Generative Pre-Training*, https://api. semanticscholar.org/CorpusID:49313245 (dostęp: 30.03.2025).

Rybak, P., Mroczkowski, R., Tracz, J., Gawlik, I. (2020). KLEJ: Comprehensive Benchmark for Polish Language Understanding. W: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (s. 1191–1201). Association for Computational Linguistics. DOI: 10.18653/v1/2020.acl-main.111.

Savchev, A. (2022). *AI Rational at CheckThat! 2022: Using Transformer Models for Tweet Classification*, "Working Notes of CLEF 2022" – Conference and Labs of the Evaluation Forum.

Sawiński, M., Węcel, K., Księżniak, E. (2024a). *OpenFact at CheckThat! 2024: Cross-Lingual Transfer Learning for Check-Worthiness Detection*, "Working Notes of CLEF 2024" – Conference and Labs of the Evaluation Forum, https://www.dei.unipd.it/~faggioli/temp/clef2024/ (dostęp: 30.03.2025).

Sawiński, M., Węcel, K., Księżniak, E. (2024b). *Under-Sampling Strategies for Better Transformer-Based Classifications Models*, "Proceedings of CLEF 2024" – Conference and Labs of the Evaluation Forum.

Sawiński, M., Węcel, K., Księżniak, E., Stróżyna, M., Lewoniewski, W., Stolarski, P., Abramowicz, W. (2023*). OpenFact at CheckThat! 2023: Head-to-Head GPT vs. BERT – A Comparative Study of Transformers Language Models for the Detection of Checkworthy Claims*, "Working Notes of CLEF 2023" – Conference and Labs of the Evaluation Forum.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I. (2017). *Attention Is All You Need*. DOI: 10.48550/arXiv.1706.03762.

8.0

HIPERPERSONALIZACJA Z UŻYCIEM AI Z ZACHOWANIEM PRYWATNOŚCI UŻYTKOWNIKA

Izabella Krzemińska
Orange Polska

Jakub Rzeźnik
Szkoła Główna Handlowa w Warszawie

Streszczenie

Rozdział przedstawia koncepcję hiperpersonalizacji w kontekście przemysłu 5.0, kluczowe różnice między dawną a nową personalizacją oraz wyzwania związane z ochroną prywatności użytkowników. Szczególną uwagę poświęcono wyjaśnieniu trzech paradoksów prywatności i personalizacji, związanych z wysokim poziomem wiedzy płynącej z danych. Wyniki naszych badań wskazują na konieczność zrównoważenia zaawansowanej personalizacji z ochroną prywatności zarówno użytkowników produktów końcowych, jak i pracowników firm korzystających z hiperpersonalizacji. Proponujemy metodę projektowania systemów AI operujących na danych użytkownika i dokonujących analiz całkowicie na urządzeniu końcowym (na przykładzie smartfonów). Przedstawiamy także przegląd technologii AI stosowanych w hiperpersonalizacji oraz metod *privacy by design* i *edge computing*.

Słowa kluczowe: personalizacja, AI, hiperpersonalizacja, przemysł 5.0, odpowiedzialna AI, prywatność, dobrostan użytkownika, zrównoważony rozwój

Wprowadzenie

W ostatnich latach koncepcja przemysłu 5.0 wprowadza nową hierarchię celów, kładąc nacisk na zrównoważony rozwój i integrację technologii z ludzkimi aspektami pracy. Celem niniejszego rozdziału jest identyfikacja roli hiperpersonalizacji w przemyśle 5.0 oraz ustalenie wyzwań związanych z ochroną prywatności użytkowników. Istnieją już rozwiązania techniczne, które sprzyjają personalizacji doświadczeń użytkownika, opierając się na danych i ochronie prawa użytkowników do zachowania prywatności.

Koncepcja przemysłu 5.0 jest nowym konstruktem, warto przywołać więc jego definicję, stworzoną jako wypadkowa dwunastu innych definicji przez Kovari [2024, s. 270]: „Przemysł 5.0 koncentruje się na efektywnej współpracy człowieka z maszynami, aby zwiększyć elastyczność i zrównoważony rozwój, polegając na inteligentnych maszynach. Przemysł 5.0 opiera się na osiągnięciach przemysłu 4.0, ale zamiast zastępować ludzi, ma na celu wykorzystanie potencjału ludzkiej inteligencji w interakcji człowiek – maszyna bardziej niż kiedykolwiek wcześniej. Pozwoli to ludziom wykorzystać swoje zdolności poznawcze i szybką adaptację do poprawy bez poświęcania dokładności i spójności oferowanych przez inteligentne maszyny. W ten sposób przemysł 5.0 koncentruje się na wpływie ekonomicznym, środowiskowym i społecznym, aby dokonywać zrównoważonych wyborów”[1].

W przemyśle 5.0 uczymy się i redefiniujemy cele na bieżąco, zmieniając akcenty i umieszczając człowieka w centrum transformacji. Głównym czynnikiem wpływającym na zmianę priorytetów jest upowszechnienie sztucznej inteligencji (*artificial intelligence*, AI) oraz pojawiające się wyzwania prawne, etyczne, społeczne. Wersja 4.0, odpowiadając na potrzeby przedsiębiorstw typu lean, takie jak eliminacja marnotrawstwa, zwiększenie efektywności, standaryzacja procesów, szybkie reagowanie na zmiany oraz ciągłe doskonalenie się, promowała automatyzację i dostosowywanie się do potrzeb użytkownika/klienta [Trzcieliński, 2020]. Z kolei przemysł 5.0 uzupełnia te potrzeby i rozszerza rozumienie człowieka w centrum zainteresowania o pracowników, podkreślając potrzebę współpracy między inteligentnymi systemami oraz zrównoważenie potrzeb ludzi (pracowników, użytkowników, konsumentów), a nie przedsiębiorstw.

---

1 Oryginalne cytowanie z artykułu w języku angielskim (tłumaczenie własne): “Industry 5.0 focuses on effective human co-work with machines to increase flexibility and sustainability by relying on smart machines. Industry 5.0 builds on the achievements of Industry 4.0, but rather than replacing humans, it aims to exploit the potential of human intelligence in human-machine interaction more than ever before. This will allow people to use their cognitive abilities and rapid adaptability to improve without sacrificing the accuracy and consistency offered by intelligent machines. In this way, Industry 5.0 focuses on economic, environmental and social impact to make sustainable choices”.

To przewartościowanie w przemyśle 5.0 odpowiada na pojawiające się wyzwania związane z integracją nowych technologii z ludzkimi aspektami pracy, zapewnieniem bezpieczeństwa i etyki danych oraz zrównoważonego rozwoju i adaptacji infrastruktury. Zakłada też nieustanną adaptację jako kluczowy czynnik zrównoważonego rozwoju, gdzie znaczącą rolę w tej przemysłowej ewolucji odgrywają coraz to nowsze technologiczne innowacje [Nahavandi, 2019; Piccarozzi, Silvestri, Silvestri, Ruggieri, 2024].

Z kolei wzrastające możliwości techniczne oraz rosnąca potrzeba personalizacji już nie tylko produktów i usług, ale też doświadczeń użytkowników sprawiają, że coraz częściej mówi się o hiperpersonalizacji. Dużym wyzwaniem staje się osiągnięcie równowagi pomiędzy personalizacją użytkownika a prywatnością. Jeszcze większym wyzwaniem jest elastyczność w sytuacji zacierających się granic wobec tego, co uznać za korzyść, a co za szkodliwość [Finck, 2021; Solove, 2020; Zuboff, 2023] wraz z rosnącą wrażliwością dotyczącą danych cyfrowych [Graepel, Kosinski, Stillwell, 2013].

Dawna personalizacja skupiała się na dostosowywaniu produktów do szerokich grup użytkowników, opierając się na analizie danych demograficznych i ogólnych charakterystykach. Współczesna hiperpersonalizacja dzięki zaawansowanym technologiom umożliwia dostarczanie spersonalizowanych rozwiązań na podstawie szczegółowych danych zebranych w czasie rzeczywistym, tworząc unikalne doświadczenia użytkowników i przewidując ich przyszłe potrzeby. Ta współczesna koncepcja znacząco różni się od dawnych metod personalizacji opartych na poszukiwaniu cech typowych dla grup, co pokazano w tabeli 8.1. Dzięki zaawansowanym technologiom, takim jak sztuczna inteligencja i big data, możliwe jest dostosowywanie produktów do indywidualnych potrzeb użytkowników w czasie rzeczywistym. Podczas gdy tradycyjna personalizacja była ograniczona do grup użytkowników o wspólnych cechach, hiperpersonalizacja koncentruje się na jednostkowych preferencjach, bazując na szczegółowej analizie danych z wielu źródeł. Teraz przestaliśmy tworzyć produkty o konkretnych cechach a zaczęliśmy tworzyć automatycznie adaptowalne doświadczenia użytkownika [Kovari, 2024; Kuksa, Skinner, Fisher, Kent, 2022]. Następuje zwrot w kierunku złożonych algorytmów, które na podstawie danych przewidują potrzeby użytkowników, wykorzystują ich zachowania, ich zaangażowanie w czasie rzeczywistym, a nawet reakcje emocjonalne, aby dostroić interakcje i oferty. Technologia umożliwia przejście od doświadczeń typowych dla segmentów do prawdziwie zindywidualizowanych, gdzie segmentów jest tyle, ilu użytkowników.

Hiperpersonalizacja usług wymaga kolejnego przesunięcia akcentów i wyznaczanych priorytetów dla firm – z wysokiej produktywności i zaawansowania technologicznego w stronę zorientowania się na dopasowanie do klientów jako domyślnej

strategii. Konsekwencją takiej postawy jest zmiana ośrodka zarządzania technologią, gdzie to docelowy użytkownik staje się jej kontrolerem i wykorzystuje narzędzia dostarczane przez przemysł 5.0 z korzyścią dla własnego dobrostanu [Alves, Gaspar, Lima, 2023]. Oprócz indywidualnych zysków to nowe podejście powinno wzmacniać zarówno rozwój ludzi, współpracę między nimi, jak i kooperację między użytkownikami a maszynami.

Tabela 8.1. Porównanie dawnej i obecnej personalizacji

Cecha	Dawna personalizacja	Nowa hiperpersonalizacja
Zasięg	oparcie segmentacji na typowych użytkownikach różnie zdefiniowanych grup np. według zmiennych demograficznych	dopasowanie do jednostkowych preferencji z wykorzystaniem dostępnych cyfrowych danych
Technologia	ograniczone narzędzia analizy danych (ankiety, badania)	zaawansowane narzędzia: AI, big data, uczenie maszynowe
Dane	dane demograficzne, historyczne, zbiorcze	dane behawioralne, w czasie rzeczywistym, wielokanałowe
Zakres personalizacji	ograniczona, powierzchowna w celu uchwycenia najbardziej typowego zakresu cech, statystyczna	głęboka, precyzyjna, indywidualna, ograniczona tylko dostępem do danych
Sposób dostosowania	reaktywne: dostosowanie na podstawie przeszłych zachowań, najczęściej zachowań dla grupy	proaktywne: przewidywanie przyszłych potrzeb z uwzględnieniem kontekstu sytuacyjnego w czasie rzeczywistym
Cele	maksymalizacja zaspokojenia potrzeb grup docelowych	tworzenie unikalnych, personalizowanych doświadczeń, adaptowanych w czasie rzeczywistym
Interakcja z użytkownikiem	rzadka, mało dynamiczna	częsta, dynamiczna, w czasie rzeczywistym
Zakres zbieranych danych	ograniczony do kilku zmiennych (wiek, płeć, użycie produktu)	wielowymiarowe dane: lokalizacja, nawyki, preferencje, dane z wymiany społecznej, poglądy
Źródła danych	ankiety, dane transakcyjne	dane z mediów społecznościowych, aplikacji, internet rzeczy (IoT)

Źródło: opracowanie własne.

8.1. Hiperpersonalizacja i paradoksy prywatności

Hiperpersonalizacja, będąca rozwinięciem personalizacji, polega na technologicznym dostosowaniu produktów i usług do indywidualnych potrzeb użytkowników na skalę większą niż kiedykolwiek wcześniej. Technologie AI, takie jak Machine Learning i analityka predykcyjna, odgrywają kluczową rolę w hiperpersonalizacji [Kuksa i in., 2022; Schrage, 2014]. Algorytmy przewidujące potrzeby użytkowników

na podstawie analizy zachowań i reakcji emocjonalnych mogą dostroić interakcje i oferty w czasie rzeczywistym.

Przykłady zastosowania tych technologii obejmują personalizowane rekomendacje produktów, dynamiczne dostosowywanie interfejsów użytkownika oraz zindywidualizowane kampanie marketingowe.

Personalizacja w cyfrowym świecie wymaga zaawansowanej analizy danych użytkowników, łącząc dostosowanie produktu, usługi doświadczeń z maksymalizacją satysfakcji użytkowników, stąd też liczne wyzwania związane z prywatnością, etyką i ochroną danych. Prowadzi też do paradoksów, trudnych do rozwiązania dotychczasowymi metodami. Obszerne publikacje autorów i autorek, takich jak Finck [2021], Solove [2020], Zuboff [2023], dotyczące pojawiających się paradoksów doprowadziły do szerokiej dyskusji i trendu poszukiwania metod, które im przeciwdziałają i wymagają omówienia w kontekście prezentowanego rozwiązania.

Hiperpersonalizacja z założenia wymaga użycia danych o użytkowniku, traktowanych często jako dane wrażliwe, co budzi obawy dotyczące ochrony prywatności. I chociaż obecne trendy w technikach personalizacji wskazują na odwrót od zasad gromadzenia dużych ilości danych (big data) w kierunku minimalizacji przetwarzania, to jednak dalej dotyczą one danych niosących za sobą duży potencjał informacji o użytkowniku lub pracowniku. Z kolei Finck [2021] pisze o paradoksie regulacyjnym. Przepisy takie jak RODO mają na celu ochronę prywatności użytkowników, ale często nie nadążają za szybkim postępem technologicznym. Technologie rozwijają się szybciej niż regulacje, które są tworzone reaktywnie, co prowadzi do luk prawnych i trudności w skutecznej ochronie ludzi. Firmy muszą więc na bieżąco dostosowywać swoje praktyki do zmieniających się przepisów, co bywa skomplikowane i kosztowne, oraz powinny angażować użytkowników w dialog dotyczący użycia ich danych [Velasco, Reinoso-Carvalho, Escobar, Gustafsson, Petit, 2024].

Trzeci paradoks stanowi możliwość późniejszego wykorzystania danych do celów, na które użytkownicy pierwotnie nie wyrazili zgody [Finck, 2021]. W miarę rozwoju technologii i pojawiania się nowych możliwości biznesowych dane zebrane w jednym kontekście mogą być używane w innym, co może naruszać zaufanie użytkowników i prowadzić do problemów prawnych. Przykładem może być użycie danych zebranych podczas zakupów online do tworzenia szczegółowych profilów konsumentów bez ich wiedzy i zgody, a używanych w dowolnym celu łącznie z ich odsprzedażą [Zuboff, 2023].

Wprowadzenie hiperpersonalizacji wiąże się także z wyzwaniami etycznymi. Firmy muszą podejmować decyzje dotyczące tego, jak dalece mogą ingerować w prywatność użytkowników w imię lepszego dostosowania usług. Istnieje ryzyko, że nadmierna personalizacja może prowadzić do uzależnień lub potencjalnej alienacji klientów, co

może negatywnie wpłynąć na ich zdrowie psychiczne i poczucie autonomii [Kuksa i in., 2022]. Dlatego kluczowym elementem zarządzania prywatnością w kontekście hiperpersonalizacji jest edukacja użytkowników. Powinni być świadomi, jakie dane są zbierane, w jaki sposób się je wykorzystuje i jakie mają prawa w zakresie ochrony swoich danych. Transparentność i jasna komunikacja ze strony firm są niezbędne do budowania zaufania i zapewnienia, że użytkownicy czują się bezpiecznie, korzystając ze spersonalizowanych usług [Pötzsch, 2009].

8.2. Poszukiwanie rozwiązania

Jedną z prób sprostania tym wyzwaniom jest koncepcja *privacy by design*, czyli prywatności uwzględnianej jako priorytet już na etapie projektowania technologii. Prywatność staje się integralnym i nienegocjowalnym elementem całego procesu tworzenia technologii, co zapewnia ochronę danych na każdym etapie ich przetwarzania [Cavoukian, 2012; Rocha, Silva, Dias, 2023]. Strategia minimalizacji danych, czyli gromadzenie tylko tych danych, które są niezbędne do realizacji określonych celów, również odgrywa kluczową rolę w ochronie prywatności użytkowników. Kolejnym filarem zwiększającym ochronę i kontrolę nad danymi jest przetwarzanie danych na urządzeniu użytkownika (*edge computing*) zamiast przesyłania ich do centralnych serwerów [Lin i in., 2024]. Minimalizuje to ryzyko wycieku danych i pozwala na szybsze przetwarzanie informacji.

Poszukiwane rozwiązanie nie mogło polegać na dotychczasowych schematach personalizowania usług, czyli rozciągniętym w czasie gromadzeniu wszelkich dostępnych informacji o zachowaniu użytkownika (np. śledzeniu odwiedzanych stron internetowych) lub przetwarzaniu i analizowaniu danych z historii używania smartfona (co ma miejsce np. w aplikacjach bankowych, sieciach społecznościowych, platformach streamingowych).

Zaproponowanym konceptem odpowiadającym na wyzwanie ochrony prywatności użytkownika w modelu *privacy by design* była idea predykcji profilu osobowości przy użyciu modelu *machine learning* już w momencie uruchomienia usługi cyfrowej, na podstawie danych dostępnych od momentu instalacji. Oprócz korzyści dla użytkowników w postaci zachowania prywatności (brak gromadzenia danych), takie rozwiązanie dawałoby przedsiębiorstwom możliwość zaoferowania personalizacji od pierwszego kontaktu z usługą. Dodatkowy benefit stanowi także możliwość użycia profilu użytkownika niezależnie od częstotliwości korzystania z usługi. Kolejnym wymaganiem dla poszukiwanego rozwiązania była pełna automatyzacja przewidy-

wania cech użytkownika, decydujących o jego potrzebach i preferencjach. Ze względu na możliwość zróżnicowanego udzielania zgód na dostęp do chronionych danych (lub brak takiej zgody) model predykcyjny musiał także brać pod uwagę wszystkie możliwe kombinacje dostępu do danych na urządzeniu końcowym.

Dodatkowo profil stworzony w wyniku chwilowego przetwarzania danych dostępnych podczas instalacji aplikacji byłby sam w sobie zminimalizowany, określając tylko te cechy, których poziom zdefiniowany jako niski (poniżej dwóch odchyleń standardowych od średniej) i wysoki (powyżej dwóch odchyleń standardowych od średniej) wskazywałby na szczególne i statystycznie różne potrzeby użytkownika. Zakładamy, że nie ma potrzeby specjalnego dostosowywania usługi dla użytkowników o przeciętnych potrzebach, zazwyczaj to właśnie dla nich tworzone i projektowane są usługi. Proponowana personalizacja skupia się na użytkownikach o potrzebach innych niż typowe, np. skrajnych introwertykach, którzy potrzebują znacznie mniej stymulacji i bodźców, oraz skrajnych ekstrawertykach, którzy potrzebują więcej stymulacji niż przeciętny użytkownik. Podobna zasada dotyczyłaby wszystkich cech, które byłyby przewidywane w modelu.

8.3. Budowanie i testowanie rozwiązania

Metodologią zastosowaną w przeprowadzeniu badania było Design Science, zaproponowane przez Hevner i Chateerjee w 2012 r. Postuluje ono postępowanie w trzech iteracyjnych cyklach: odpowiedniości, rygoru i projektowania artefaktów [Peffers i in., 2007]. Pierwszym krokiem jest identyfikacja problemu i motywacji do przeprowadzenia badania nad rozwiązaniem. W kolejnym etapie weryfikuje się potencjalną użyteczność artefaktów w stosunku do istniejącej bazy wiedzy. Następnie definiuje się cele dla rozwiązania i projektuje oraz rozwija docelowe rozwiązanie. Kolejnym krokiem jest ocena artefaktów, mierząca ich efektywność, a następnie ostateczne wyniki są komunikowane właściwym interesariuszom [Venable, Pries-Veje, Baskerville, 2016].

Proces badawczy rozpoczął się od wyboru koncepcji personalizacji i oparciu jej na cechach osobowości użytkowników. W obszarach badań związanych ze śladami cyfrowymi popularnym i wielokrotnie zweryfikowanym modelem osobowości jest *big 5* [Krzemińska, Rzeźnik, 2021]. Pięć wielkich czynników osobowości to: otwartość na doświadczenie (tolerancja wobec nowego i nieznanego), sumienność (znoszenie niepewności i chaosu), ekstrawersja (tolerancja na wysoką ilość bodźców i potrzeba socjalizacji), ugodowość (skupienie na potrzebach innych i chęć współpracy), stabilność emocjonalna (znoszenie stresu) [Costa, McRae, 1992].

Kolejnym etapem było przeprowadzenie badań jakościowych, weryfikujących zróżnicowanie korzystania z usług cyfrowych oraz identyfikacja potrzeb użytkowników w zależności od profilu osobowości, wyznaczanego kwestionariuszem IPIP-50. Następnie w standardowej procedurze psychometrycznej utworzono własne narzędzie do przeprowadzenia testu osobowości. W dalszej kolejności powstała aplikacja na smartfony z systemem Android, w której połączone zostały dwie funkcje – wypełnianie testu osobowości oraz zbieranie i przetwarzanie danych użytkowników aplikacji. Kategorie analizowanych danych to: dane telekomunikacyjne (np. historia połączeń telefonicznych, wysyłania wiadomości tekstowych), ustawienia telefonu (np. jasność ekranu, wielkość tekstu na ekranie, poziom baterii), dane statystyczne dotyczące zdjęć (np. kolorystyka, liczba wykrytych twarzy, tryb czarno-biały) oraz dane dotyczące zainstalowanych aplikacji (np. lista aplikacji, częstość ich użycia).

Tabela 8.2. Miary jakości modelu predykcji osobowości uzyskane w badaniu

Cecha osobowości	Dokładność (*accuracy*)	F1
Otwartość	0,73	0,68
Sumienność	0,69	0,62
Ekstrawersja	0,76	0,73
Ugodowość	0,72	0,63
Stabilność emocjonalna	0,69	0,62

Źródło: opracowanie własne.

Pozyskiwanie i analizowanie danych zostało przeprowadzone na ponad 5000 uczestników płatnego panelu badawczego, którzy zostali poinformowani o zbieraniu anonimowych danych (bez łączenia z jakimikolwiek identyfikatorami) oraz udzielili niezbędnych zgód na dostęp do danych przed rozpoczęciem korzystania z aplikacji. Ostatecznie zebrano ponad 200 zmiennych różnego rodzaju, które po połączeniu z wynikiem testu osobowości zostały na późniejszym etapie wykorzystane w tworzeniu modelu predykcyjnego. Model machine learning przewidujący profil osobowości został wytrenowany przy użyciu bibliotek *scikit-learn*, *pandas*, *numpy*. Ewaluacja jakości modelu nastąpiła poprzez użycie metryk *accuracy*, *precision* i *F1 score* na zbiorze testowym oraz poprzez walidację krzyżową. Wyniki, przedstawione w tabeli 8.2, pozwalają na stwierdzenie, że stworzony przez uczenie maszynowe model, przewidujący profil osobowości, osiągnął trafność (*accuracy*) 0,76 dla ekstrawersji. To wskazuje na wysoką skuteczność modelu w przewidywaniu potrzeb użytkowników na podstawie danych z ich urządzeń mobilnych. Nieco niższe wyni-

ki osiągnięto dla pozostałych cech, w tym najniższe dla sumienności i stabilności emocjonalnej (0,69 i 0,62 odpowiednio). Oznacza to, że model przewidywania osobowości wypracowany w badaniu osiągnął lepsze wyniki niż benchmark utworzony z 14 badań predykcji ze *state-of-the-art* (gdzie średnia dla cech z najwyższym *accuracy* wyniosła 0,69, a średnia dla wszystkich cech: 0,643), ustępując jedynie modelowi korzystającemu ze śledzenia ruchów gałek ocznych (tu osiągnięto accuracy 0,85), jak wykazuje Krzemińska [2022]. Procedura tworzenia modelu według metodologii Design Science została szczegółowo opisana w pracy doktorskiej autorki niniejszego rozdziału [Krzemińska, 2022].

Po wytrenowaniu i walidacji modelu predykcyjnego, działającego na danych ze smartfonów, możliwe stało się utworzenie osobnego artefaktu – biblioteki systemu Android. Dzięki temu w każdą aplikację na telefony mobilne z powyższym systemem może zostać wbudowany komponent przewidujący profil klienta. Jest to możliwe dzięki wytrenowaniu modelu predykcyjnego na danych dostępnych od momentu instalacji oraz połączeniu ich z wynikami testu osobowości. W przypadku nowych klientów model oczywiście nie zna profilu ich osobowości, ale jest w stanie go przewidywać na podstawie swojego treningu. Testy techniczne wykazały, że czas przetwarzania danych lokalnie i określenia osobowości klienta wynosi poniżej jednej sekundy. Aplikacje korzystające z tego typu personalizacji mogą więc dopasować swój wygląd i funkcjonalności do użytkownika już w trakcie pierwszego uruchamiania. Unikalną cechą tego rozwiązania jest to, że wszystkie obliczenia odbywają się na urządzeniu użytkownika, a dane nie są przesyłane na zewnątrz, co chroni prywatność. Sugeruje to rozwiązanie dla trzech przywoływanych wcześniej paradoksów prywatności:

1. Pożądane przez użytkowników dostosowanie usługi do ich indywidualnych potrzeb ma miejsce na urządzeniu końcowym (w tym przypadku na smartfonie) na podstawie jednorazowego, lokalnego dostępu aplikacji, na który wyrażają zgodę.
2. Nawet w sytuacji, gdy regulacje prawne nie nadążają za postępem technologicznym, przedsiębiorstwo dostarczające aplikację z komponentem personalizującym w sposób zachowujący prywatność nie ma dostępu do danych, dzięki którym dokonywana jest indywidualizacja produktu. Nie ma więc możliwości wnioskowania czegoś o kliencie z danych, które nie są objęte regulacjami.
3. W proponowanym rozwiązaniu nie istnieje przechowywanie danych, nie ma zbierania historii działań użytkownika. Oznacza to, że dane nie mogą zostać użyte w późniejszym terminie do zastosowań, na które użytkownicy nie wyrazili zgody.

8.4. Dyskusja

Era paradoksów prywatności wymaga szczególnego zrozumienia potencjału i pułapek technologii personalizacji. Nasze badania wskazują na konieczność dalszej eksploracji potencjału *privacy by design* oraz *edge computing* w kontekście hiperpersonalizacji. Przyszłe badania powinny skoncentrować się na rozwijaniu narzędzi AI, które łączą zaawansowaną personalizację z ochroną prywatności, oraz na analizie długoterminowych skutków takich technologii dla użytkowników. Spersonalizowane doświadczenia mogą znacznie zwiększyć zaangażowanie i satysfakcję użytkowników, ale nie powinny odbywać się kosztem naruszenia prywatności. Zrównoważenie tych aspektów wymaga solidnych rozwiązań technologicznych, rygorystycznych przepisów oraz zmiany podejścia do danych użytkowników i w konsekwencji zmiany w podejściu biznesowym. Poszanowanie prywatności i autonomii użytkowników w przemyśle 5.0 powinno być priorytetem.

Rozwiązania podobne do zaproponowanych w rozdziale nie mogą być wdrażane bez uprzedniej dokładnej i uczciwej analizy korzyści użytkowników i biznesu. Świadomość zagrożeń pojawiających się przy stosowaniu personalizacji [Zuboff, 2023] powinna doprowadzić do uzupełnienia technologii o mechanizmy chroniące użytkownika, np. zapobiegające uzależnieniu lub nieświadomej utracie kontroli i autonomii.

Niewątpliwie dalszych skrupulatnych i pogłębionych badań wymagają zarówno biznesowe ograniczenia personalizacji, jak i te związane z ograniczeniami samego użytkownika. Myśląc o ograniczeniach biznesowych, można stawiać pytania: Czy personalizacja ma granice? Czy personalizując w sposób prawie doskonały i zindywidualizowany wszystkie usługi cyfrowe, doprowadzimy do sytuacji, w której personalizacja wyczerpie swój korzystny potencjał? Czy istnieją jakieś alternatywy? Z kolei myśląc o ograniczeniach ze strony użytkownika, należy rozważyć i dobrze przebadać aspekty związane z różnym poziomem potrzeb w zakresie personalizacji, poczucia znudzenia lub zagrożenia nadmierną ilością wiedzy o użytkowniku lub takim poziomem personalizacji, który zaczyna użytkownikowi zagrażać (uzależnienie, zbyt duża atrakcyjność cyfrowych doświadczeń). Oddzielnym aspektem wymagającym dalszych badań jest ryzyko braku rozwoju poznawczego i emocjonalnego w przypadku nadmiernego dostosowywania się lub nawet ryzyka braku potrzeby socjalizacji – bo człowiek, mając własne autonomiczne cele, nie osiągnie takiego poziomu adaptywności, jak potencjalnie może osiągnąć model oparty na sztucznej inteligencji. To ostatnie z kolei jest wyzwaniem społecznym, bo sprzyja rosnącej samotności.

Prezentowane rozwiązanie, choć realizuje cele związane z prywatnością oraz oszczędnym korzystaniem z zasobów (minimalizacja ilości danych i przetwarzania),

wymaga rozwiązań uzupełniających, związanych ze skalowalnością oraz opracowaniem mierników jakości danych do stworzenia modelu oraz jakości samego uzyskanego profilu. Powinno być wzbogacone o ocenę zaufania do obu tych czynników. Stanowi to poważne wyzwanie w sytuacji braku gromadzenia danych oraz zachowania prywatności danych. Istnieje już szereg prac badawczych w obszarze *federated analytics* i *federated learning*, których celem jest budowa rozwiązań dedykowanych wymianie niezbędnych informacji przy zachowaniu prywatności [Liu i in., 2024].

Podsumowanie

W kontekście przemysłu 5.0 kluczowym wyzwaniem jest znalezienie równowagi między zaawansowaną personalizacją a ochroną prywatności użytkowników. Nasze badania wskazują, że zastosowanie koncepcji *privacy by design* oraz strategii minimalizacji danych może skutecznie adresować te wyzwania. Proponowane rozwiązanie, polegające na lokalnym przetwarzaniu danych, minimalizacji zbieranych informacji oraz integracji ochrony prywatności już na etapie projektowania, zapewnia wyższą jakość spersonalizowanych usług oraz gwarantuje bezpieczeństwo danych użytkowników.

Jednym z przyszłych kierunków rozwoju technologii personalizacji w przemyśle 5.0 jest wykorzystanie *federated learning*. Podejście to pozwala na trenowanie modeli AI bez konieczności centralizowania danych użytkowników. Dzięki temu modele mogą być aktualizowane na podstawie danych lokalnych, bez gromadzenia danych trenujących, co zwiększa poziom ochrony prywatności, minimalizuje ryzyko wycieku informacji oraz pozwala na ciągłe doskonalenie algorytmów.

Dodatkowo *federated analytics* może być stosowane do monitorowania aktualności wykorzystywanych modeli bez narażania prywatności użytkowników. Dzięki temu firmy mogą na bieżąco oceniać efektywność personalizacji i wprowadzać niezbędne poprawki, co zapewnia utrzymanie usług wysokiej jakości.

Przyszłe badania powinny koncentrować się na dalszym rozwijaniu technologii *federated learning* i *federated analytics*, a także na analizie ich wpływu na ochronę prywatności i jakość spersonalizowanych usług. Dalszy rozwój tych technologii, w połączeniu z edukacją użytkowników w zakresie ochrony ich danych, będzie kluczowy dla sukcesu przemysłu 5.0 i zapewnienia zrównoważonego rozwoju technologii personalizacji.

W przyszłości dalszy rozwój technologii przemysłu 5.0 powinien koncentrować się na edukacji użytkowników (komunikacja korzyści i zagrożeń związanych z personalizacją), rozwoju technologii sprzyjających ochronie prywatności oraz harmonizacji

technologicznych innowacji z wartościami społecznymi i etycznymi. Podobne rozwiązania będą stanowiły istotne wsparcie dla realizacji celów przemysłu 5.0.

Prezentowane rozwiązanie zapewnia nie tylko wyższą jakość spersonalizowanych usług, ale również gwarantuje bezpieczeństwo i ochronę danych użytkowników, co jest kluczowe w dynamicznie zmieniającym się środowisku przemysłowym.

Należy przypuszczać, że przewagę konkurencyjną zdobędzie ten, kto zrozumie, że można dostarczyć użytkownikowi zindywidualizowane doświadczenie z poszanowaniem jego praw. Firmy muszą oduczyć się pokusy gromadzenia informacji o kliencie i skupić się na tworzeniu wartości poprzez ochronę prywatności. Integracja nowych technologii z ludzkimi aspektami pracy oraz zapewnienie bezpieczeństwa i etyki danych będą kluczowe dla sukcesu przemysłu 5.0.

Bibliografia

Alves, J., Gaspar, P.D., Lima, T.M. (2023). Is Industry 5.0 a Human-Centered Approach? A Systematic Review, *Processes*, 11(1), s. 193.

Cavoukian, A. (2012). Privacy by Design [Leading Edge], *IEEE Technology and Society Magazine*, 31(4), s. 18–19.

Costa Jr, P.T., McCrae, R.R. (1992). Four Ways Five Factors Are Basic, *Personality and Individual Differences*, 13, s. 653–665.

Finck, M. (2021). The Limits of the GDPR in the Personalisation Context. W: *Data-Driven Personalisation in Markets, Politics and Law* (s. 95–107), U. Kohl, J. Eisler (Eds). Cambridge: Cambridge University Press.

Kiepas, A. (2020). Człowiek w świecie procesów cyfryzacji – współczesne wyzwania i przyszłe skutki, *Filozofia i Nauka. Studia Filozoficzne i Interdyscyplinarne*, 8(1), s. 155–174.

Kosinski, M., Stillwell, D., Graepel, T. (2013). Private Traits and Attributes Are Predictable from Digital Records of Human Behavior, *PNAS Proceedings of the National Academy of Sciences of the United States of America*, 110(15), s. 5802–5805. DOI: 10.1073/pnas.1218772110.

Kovari, A. (2024). Industry 5.0: Generalized Definition, Key Applications, Opportunities and Threats, *Acta Polytechnica Hungarica*, 21(3), s. 267–284.

Krzemińska, I. (2022). *Automatic Data-Based Personality Assessment As a Method of Electronic Services Auto-Personalisation (PhD Thesis)*, https://bip.ue.poznan.pl/download/attachment/1325/rozprawa-doktorska-mgr-i-krzeminskiej.pdf (dostęp: 14.10.2024).

Krzemińska, I., Rzeźnik J. (2021). Personality-Based Lexical Differences in Services Adaptation Process, *Technium Romanian Journal of Applied Sciences and Technology*, 3, s. 61–73.

Kuksa, I., Skinner, M., Fisher, T., Kent, T. (2022). *Delivering Personalised, Digital Experience*. W: *Understanding Personalisation: New Aspects of Design and Consumption* (s. 67–87). Amsterdam: Elsevier Science & Technology.

Lin, Q., Jiang, S., Zhen, Z., Chen, T., Wei, C., Lin, H. (2024). Fed-PEMC: A Privacy-Enhanced Federated Deep Learning Algorithm for Consumer Electronics in Mobile Edge Computing, *IEEE Transactions on Consumer Electronics*, 70(1), s. 4073–4086. DOI: 10.1109/TCE.2024.3351648.

Liu, Y., Kang, Y., Zou, T., Pu, Y., He, Y., Ye, X., Ouyang, Y., Zhang, Y.Q., Yang, Q. (2024). Vertical Federated Learning: Concepts, Advances, and Challenges, *IEEE Transactions on Knowledge and Data Engineering*, 36(7), s. 3615–3634. DOI: 10.1109/TKDE.2024.3352628.

Nahavandi, S. (2019). Industry 5.0 – A Human-Centric Solution, *Sustainability*, 11(16), 4371.

Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research, *Journal of Management Information Systems*, 24(3), s. 45–77.

Rocha, L.D., Silva, G.R.S., Canedo, E.D. (2023). Privacy Compliance in Software Development: A Guide to Implementing the LGPD Principles. W: *Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing* (s. 1352–1361). DOI: 10.1145/3555776.3577615.

Piccarozzi, M., Silvestri, L., Silvestri, C., Ruggieri, A. (2024). Roadmap to Industry 5.0: Enabling Technologies, Challenges, and Opportunities Towards a Holistic Definition in Management Studies, *Technological Forecasting and Social Change*, 205, 123467.

Schrage, M. (2014). *The Innovator's Hypothesis: How Cheap Experiments Are Worth More Than Good Ideas*. Cambridge: The MIT Press.

Trzcieliński, S. (2003). Lean Management a wirtualność przedsiębiorstwa, *Prace Naukowe Instytutu Organizacji i Zarządzania Politechniki Wrocławskiej. Konferencje (23)*, 73, s. 291–306.

Velasco, C., Reinoso-Carvalho, F., Escobar, F.B., Gustafsson, A., Petit, O. (2024). Paradoxes, Challenges, and Opportunities in the Context of Ethical Customer Experience Management, *Psychology and Marketing*. DOI: 10.1002/mar.22069.

Venable, J., Pries-Heje, J., Baskerville, R. (2016). FEDS: a Framework for Evaluation in Design Science Research, *European Journal of Information Systems*, 25, s. 77–89.

Zuboff, S. (2023). The Age of Surveillance Capitalism. W: *Social Theory Re-Wired* (s. 203–213), W. Longhofer, D. Winchester (Eds.). New York: Routledge.

9.0

TRANSFORMACJA DANYCH W CZASIE RZECZYWISTYM: OCENA WYDAJNOŚCI APACHE SPARK

Mariusz Rafało
Szkoła Główna Handlowa w Warszawie
Emilia Kaleta-Pazdur
Autorka niezrzeszona

Streszczenie

Apache Spark to platforma do przetwarzania danych w trybie wsadowym oraz w czasie rzeczywistym. Ważne wyzwanie, zwłaszcza w trybie przesyłania w czasie rzeczywistym, stanowi zapewnienie wysokiej przepustowości i niskiego opóźnienia. Jest to szczególnie ważne w przypadku systemów AI działających w czasie rzeczywistym. W niniejszym rozdziale przedstawiamy serię eksperymentów mających na celu identyfikację parametrów i konfiguracji Apache Spark, które mogą zmniejszyć opóźnienia w transformacji danych. W badaniu zweryfikowaliśmy najpopularniejsze transformacje danych: grupowanie i filtrowanie. Kluczowe ustalenia wskazują, że wybór klastra jednowęzłowego może zapewnić korzystny kompromis i osiągnięcie równowagi między wydajnością przesyłania strumieniowego a efektywnością wykorzystania zasobów.

Słowa kluczowe: Apache Spark, wydajność przetwarzania, dane w czasie rzeczywistym

Wprowadzenie

Dane w czasie rzeczywistym odnoszą się do informacji, które są gromadzone, przetwarzane i wykorzystywane natychmiast po ich wygenerowaniu, umożliwiając działanie w oparciu na najnowszych informacjach. Technologia Apache Spark jest obecnie standardem w zakresie masowego przetwarzania danych. To platforma open source do równoległego przetwarzania danych w pamięci operacyjnej [Apache Spark, 2020b]. Apache Spark jest także używany w usługach wiodących dostawców chmurowych, takich jak Amazon EMR [Amazon, 2019], Databricks [Databricks, 2023] lub Google Dataproc [Google Cloud, 2022]. Spark oferuje ujednolicony system do obsługi dużych zbiorów danych, algorytmów uczenia maszynowego i możliwość przetwarzania danych w czasie rzeczywistym. Działa na jednym serwerze lub w klastrach złożonych z wielu węzłów (serwerów). Moduł Spark Streaming jest komponentem platformy Spark, służącym do przetwarzania danych w czasie rzeczywistym.

Inną technologią, powszechnie używaną do przesyłania strumieniowego danych, jest Apache Kafka. Podczas gdy Spark Streaming oferuje przetwarzanie danych i możliwości analityczne, Apache Kafka jest rozproszoną platformą przesyłania strumieniowego. Integracja obu platform upraszcza architekturę dla analiz w czasie rzeczywistym [Salloum, Dautov, Chen, Peng, Huang, 2016].

Spark działa w rozproszonym środowisku, w którym dane i zadania obliczeniowe są rozproszone na wielu serwerach [serwery w klastrze nazywa się potocznie węzłami (*nodes*)]. Architektura ta powoduje wyzwania w diagnozowaniu wąskich gardeł wydajności, ponieważ problemy mogą wynikać z różnych przyczyn, takich jak opóźnienie sieci, nierównomierna dystrybucja danych lub nieefektywna alokacja zasobów. Ponadto w Spark Streaming wyzwania te zostają spotęgowane przez potrzebę niskich opóźnień transformacji danych [Karau, Warren, 2017]. Zarządzanie przetwarzaniem w trybie strumieniowym wymaga utrzymywania i aktualizowania stanu danych w różnych partiach danych oraz różnych punktach czasu. Ponadto optymalizacja wykorzystania zasobów (takich jak pamięć i procesor) staje się bardziej złożona, ponieważ system musi znaleźć równowagę między przetwarzaniem o niskim opóźnieniu a ograniczeniami zasobów.

Co więcej, narzędzia diagnostyczne i metryki do monitorowania wydajności w Spark wymagają zrozumienia wewnętrznej architektury platformy. Przykładowo, prezentowane w Spark parametry diagnostyczne, takie jak wykorzystanie pamięci czy procesorów, nie dają jasnej wiedzy, w jaki sposób dane odczyty przekładają się na opóźnienia. Ten poziom wiedzy specjalistycznej nie zawsze jest łatwo dostępny.

Czynniki te sprawiają, że zarządzanie wydajnością w Spark, szczególnie w module Spark Streaming, to wymagające i złożone zadanie.

Celem rozdziału jest identyfikacja możliwości zwiększenia wydajności platformy Apache Spark w wybranych scenariuszach transformacji danych w czasie rzeczywistym. Przetwarzanie danych w czasie rzeczywistym ma duże znaczenie dla systemów sztucznej inteligencji, ponieważ modele AI mogą wówczas budować swoje predykcje w oparciu na najnowszych informacjach. Udostępnianie do modeli AI danych w czasie rzeczywistym pozwala na ciągłe douczanie modeli, uwzględniając nowe zdarzenia.

W badaniu skupiamy się wyłącznie na module Spark Streaming. Ponadto nie badamy wyłącznie technologii Spark, ale także jej integrację ze strumieniami Apache Kafka; kontrolujemy szybkość przepływu danych, co pozwala zweryfikować, w jaki sposób parametry Spark działają dla różnych szybkości transmisji danych.

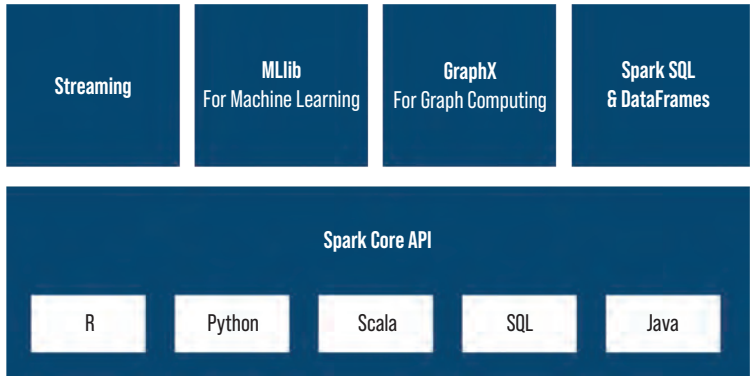
9.1. Technologia Apache Spark

Zwykle klaster Apache Spark składa się z serwera głównego (węzła) i wielu węzłów roboczych. W klastrze działają dwa typy procesów: sterownik (*driver*) i wykonawca (*executor*). Proces sterownika to główna jednostka przetwarzania danych; działa na węźle głównym i odpowiada za pozyskiwanie zasobów sprzętowych z klastra. Proces sterownika rozpoczyna się od przetłumaczenia aplikacji (kodu źródłowego) na zadania Spark. Następnie każde zadanie jest tłumaczone na logiczny plan wykonania, reprezentowany jako skierowany graf acykliczny (DAG). Wyjątek stanowią klastry jednowęzłowe, w których zarówno proces sterownika, jak i proces wykonawczy są uruchamiane na serwerze głównym. Każdy wykonawca ma wyłączne zasoby sprzętowe do obsługi zadań przetwarzania [Chao, Shi, Gao, Luo, Wang, 2018].

Transformacje danych w Apache Spark to procesy, które przekształcają zbiór danych z jednej formy w drugą. Reprezentacją struktury danych są w Spark tzw. ramki danych (*data frame*), które przypominają swoją budową tabelę (posiadają kolumny o określonych typach, zaś grupy kolumn są zorganizowane w wierszach).

Spark oferuje możliwość stosowania powszechnie używanych języków programowania, takich jak: Java, Python, SQL, R i Scala, oraz przetwarzania danych z różnych źródeł i w różnych formatach, takich jak płaskie pliki CSV, Parquet, Avro, ORC, JSON itp. Posiada kilka wbudowanych modułów (poza Spark Streaming), które służą do przetwarzania danych w zależności od celu, takich jak: Spark SQL przetwarzający dane strukturalne, MLib – biblioteka uczenia maszynowego czy GraphX przetwarzający grafy (rysunek 9.1).

Rysunek 9.1. Komponenty Apache Spark



Źródło: Damji, Wenig, Das, Lee [2020].

Informacje dotyczące głównych cech Apache Spark zostały zawarte w następujących stwierdzeniach [Apache Spark, 2020a]:

1. Dane w Spark są partycjonowane przy użyciu tzw. partycji haszujących. Partycjonowanie haszujące próbuje równomiernie rozłożyć dane na podstawie klucza [Chambers, Zaharia, 2018; Damji *et al.*, 2020].
2. Spark buforuje dane w pamięci operacyjnej na wielu węzłach.
3. Dzięki zastosowaniu ramek danych Spark zapewnia wysoką dostępność w przypadku awarii. Jeśli dany węzeł ulegnie awarii, Spark może ponownie obliczyć utraconą partycję z oryginalnego zestawu danych.
4. Spark jest wysoce skalowalny, z możliwością obsługi petabajtów (milionów gigabajtów) danych na wielu węzłach.

Zarządzanie wydajnością klastra Spark jest złożone ze względu na rozproszoną naturę Spark i potrzebę optymalizacji na różnych warstwach, w tym partycjonowania danych, serializacji i zarządzania pamięcią. Każda warstwa ma własny zestaw parametrów, które należy dostroić pod kątem zgodności z innymi. Spark oferuje ponad 180 parametrów konfiguracyjnych [Apache Spark, 2020a], z których kilkadziesiąt dotyczy dostrajania wydajności.

9.2. Badania powiązane

Badania wydajności Apache Spark skupiają się na dwóch strumieniach badawczych: ocenie wydajności [Wang, Khan, 2015] i przewidywaniu (predykcji) wydajności [Cheng, Ying, Wang, 2021]. Strumienie te jednak koncentrują się głównie na prze-

twarzaniu wsadowym. Strumień badania wydajności Spark dotyczy porównywania wydajności technologii przy wykorzystaniu testów porównawczych i dostrajania parametrów Spark. Popularnymi scenariuszami testów porównawczych wydajności są WordCount i Terasort [IBM, 2021]. Na przykład Ahmed, Barczak, Rashid i Susnjak [2021] wykonali testy oparte na tych scenariuszach dla 18 parametrów Spark. Kluczowym odkryciem było to, że Spark ma wyższą wydajność w porównaniu z technologią Hadoop MapReduce. Spark okazał się szybszy dwukrotnie w teście WordCount i czternastokrotnie w przypadku TeraSort. Naukowcy wykorzystują również kompleksową platformę testową HiBench [Ahmed *et al.*, 2021] – zunifikowaną do oceny wydajności systemów intensywnie wykorzystujących dane. Samadi, Zbakh i Tadonki [2018] przeprowadzili porównanie technologii Hadoop MapReduce z technologią Spark, korzystając z tej platformy. Okazało się, że Spark odznacza się większym wykorzystaniem procesora i operacji wejścia/wyjścia, a jednocześnie wyższą wydajnością niż MapReduce (szczególnie w przypadku obsługi większej ilości danych). W tym nurcie badawczym istnieją również badania bezpośrednio skupiające się na wydajności algorytmów AI działających na Spark [Mavridis, Karatza, 2017].

Nurt badawczy związany z predykcją wydajności Spark jest np. reprezentowany przez Petridisa, Gounarisa i Torresa [2017]. Badacze wskazują, w jaki sposób zastąpić domyślne wartości parametrów wartościami oszacowanymi przez modele predykcyjne. Badanie koncentrowało się głównie na algorytmach sortowania, kompresji i zarządzaniu pamięcią. Dowiedziono w nim, iż największy wpływ na jakość predykcji wydajności Apache Spark ma parametr *spark.shuffle.compress*.

Inne podejście do predykcji i dostrajania parametrów związanych z wydajnością zostało wprowadzone przez Petrova, Butakova, Nasonova i Melnika [2018]. W artykule zaproponowano adaptacyjny model wydajności, dedykowany skalowaniu zasobów systemowych w zależności od wymagań dotyczących opóźnień. Przeprowadzono zestaw eksperymentów, który pozwolił na uzyskanie 66% poprawy wydajności.

9.3. Zakres eksperymentu

Eksperyment polegał na uruchomieniu serii testów przetwarzania Spark Streaming zintegrowanych z Apache Kafka. Zadania Spark obejmowały połączenie z określonym strumieniem Kafka – tematem (*topic*), a następnie zebranie danych i przeprowadzenie transformacji. Eksperyment został przygotowany w języku programowania Python. Kody źródłowe umieszczono w publicznym repozytorium Github. Utworzono dwa programy: program producenta, który wysyłał dane do tematu Kafka z określoną

prędkością, oraz program konsumenta, który odczytywał dane z tematu, wykonywał odpowiednie transformacje i zapisywał wyniki. Przygotowaliśmy szereg scenariuszy testowych – każdy z nich obejmuje połączenie Apache Spark z tematem Kafka, pobranie przychodzących danych i ich transformację. Scenariusze zostały zdefiniowane przez:

- konfigurację prędkości transmisji danych (600, 1200, 1800 i 3000 rekordów na minutę);
- konfigurację liczby partycji w temacie Kafka (1, 2 lub 3);
- konfigurację sprzętową klastra Apache Spark (liczba węzłów);
- typ transformacji danych (filtrowanie i grupowanie);
- ustawienia parametrów Spark.

Wszystkie scenariusze testowe zostały uruchomione w infrastrukturze usług Amazon EC2. Platforma Apache Spark została dostarczona przez środowisko Databricks. Wykorzystaliśmy Structured Streaming API w Apache Spark. Każdy z wykonawców w klastrach miał 4 rdzenie procesora i 30,5 GB pamięci RAM.

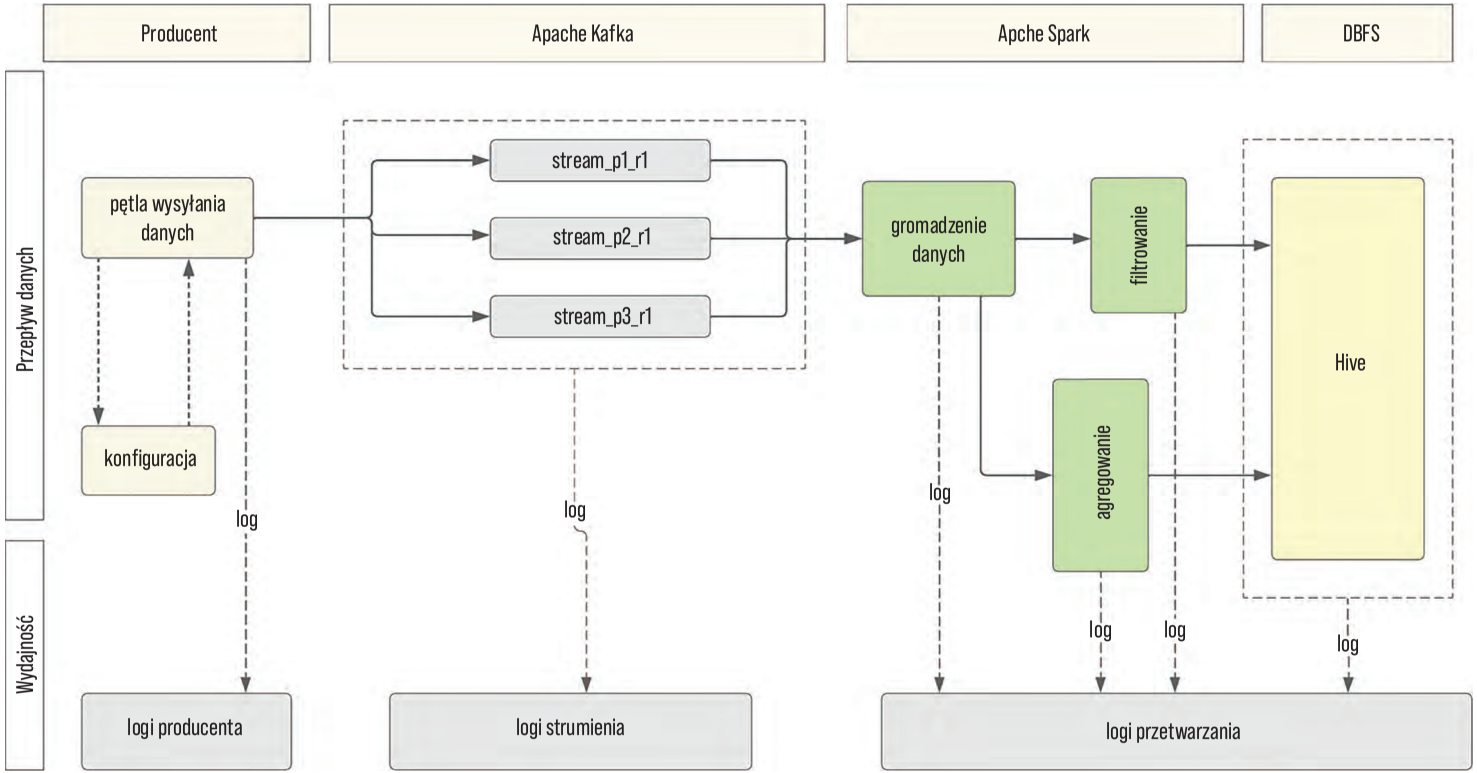
9.3.1. Przepływ danych

Koncepcję eksperymentu i przepływ danych przedstawiono na rysunku 9.2. Producent generuje dane w pętli, na podstawie konfiguracji zapisanej w pliku. Po stronie Spark konsument łączy się online ze strumieniem danych, po czym odczytuje dane z określonego tematu bezpośrednio do ramki danych. Następnie dane są transformowane.

Na każdym etapie transmisji i transformacji danych konsument rejestruje informacje, zawierające znacznik czasu na danym etapie. Posiadając dane znacznika czasu na każdym etapie przetwarzania danych, może zaś obliczyć opóźnienia (latencje) dla każdego kroku. Następnie obliczone opóźnienie jest zapisywane w repozytorium plików, wraz z konfiguracją Spark i Kafka oraz stanem infrastruktury sprzętowej klastra Spark.

Podczas badania przeanalizowaliśmy wszystkie parametry Apache Spark, aby wybrać te, które (zgodnie z dokumentacją) mogą mieć wpływ na wydajność strumienia danych. Następnie podczas oceny zidentyfikowaliśmy parametry, mające największy wpływ na wydajność przetwarzania danych.

Rysunek 9.2 Koncepcja przepływu danych w eksperymentach



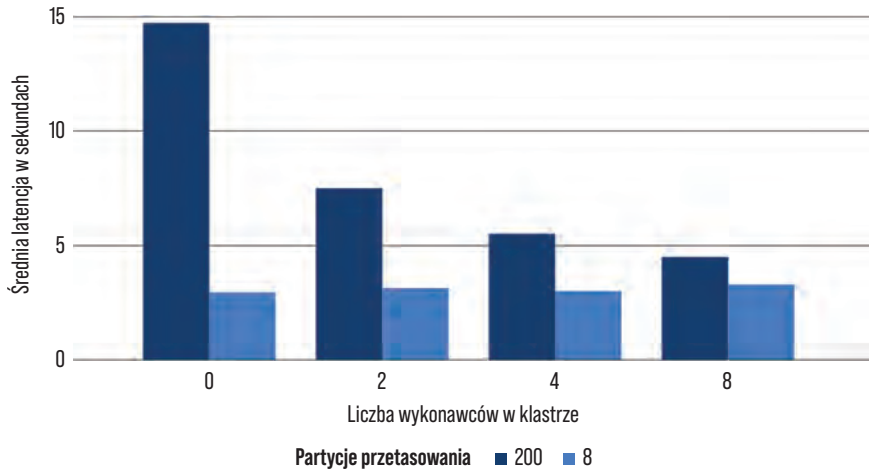
Źródło: opracowanie własne.

9.3.2. Ocena eksperymentu

Eksperyment został przeprowadzony przy stałej prędkości danych podczas danego testu. Z tego powodu utrzymaliśmy domyślną konfigurację Spark dla parametrów: maxOffsetsPerTrigger i minOffsetsPerTrigger, które w obu przypadkach były ustawione na wartość none. Wskaźnikiem wydajności była najniższa latencja, ale braliśmy również pod uwagę efektywne wykorzystanie zasobów.

Skonfigurowaliśmy skrypt inicjalizacyjny, który kontroluje częstotliwość zbierania logów. Następnie w aplikacji Spark ustawiliśmy parametr konfiguracyjny: spark.sql.streaming.metricsEnabled na true, aby włączyć gromadzenie logów. Logi metryk były zapisywane w pięciosekundowych odstępach. Biorąc pod uwagę wszystkie kombinacje scenariuszy, warianty prędkości transmisji i parametry, obliczyliśmy, że cały eksperyment trwał ponad 40 godzin.

Rysunek 9.3. Średnia latencja dla grupowania przy różnych ustawieniach parametru liczby partycji przetasowania danych



Źródło: opracowanie własne.

Rysunek 9.3 przedstawia wyniki dla scenariuszy z grupowaniem z różną konfiguracją parametru liczby partycji przetasowania danych (*shuffle partitions*): 200 (domyślnie) lub 8. Przetasowanie danych jest istotne dla transformacji takich jak grupowanie i oznacza, że uzyskanie ostatecznego wyniku wymaga przemieszania ich między partycjami Spark. Natomiast transformacje takie jak filtrowanie są wykonywane bez przetasowania danych, dlatego zmiana konfiguracji tego parametru nie wpływa na testy z filtrowaniem. Wykres pokazuje znaczący spadek latencji przy zmniejszonej liczbie

partycji przetasowania danych. Spadki latencji stają się mniejsze wraz ze wzrostem liczby wykonawców w klastrze. Niemniej jednak najniższa średnia latencja występuje dla klastra jednowęzłowego przy ustawieniu ośmiu partycji przetasowania danych. Dodatkowo poddaliśmy analizie wielkość próby i odchylenie standardowe dla scenariuszy przedstawionych na rysunku 9.3. Dane zebrane w tabeli 9.1 pokazują relatywnie duże liczności prób, przy przeważająco niskich odchyleniach standardowych. Biorąc pod uwagę tę analizę, wykluczyliśmy z dalszych testów dane oparte na domyślnej konfiguracji liczby partycji przetasowania danych (200) ze względu na wyższe odchylenie standardowe.

Tabela 9.1. Liczebność prób i odchylenie standardowe dla danych zaprezentowanych na rysunku 9.3

Liczba wykonawców	Partycje przetasowania	Liczebność próby	Odchylenie standardowe
0	8	2123	1,40
0	200	3463	3,66
2	8	2132	0,52
2	200	2046	2,25
4	8	2309	0,82
4	200	2266	1,91
8	8	2128	0,42
8	200	2104	0,84

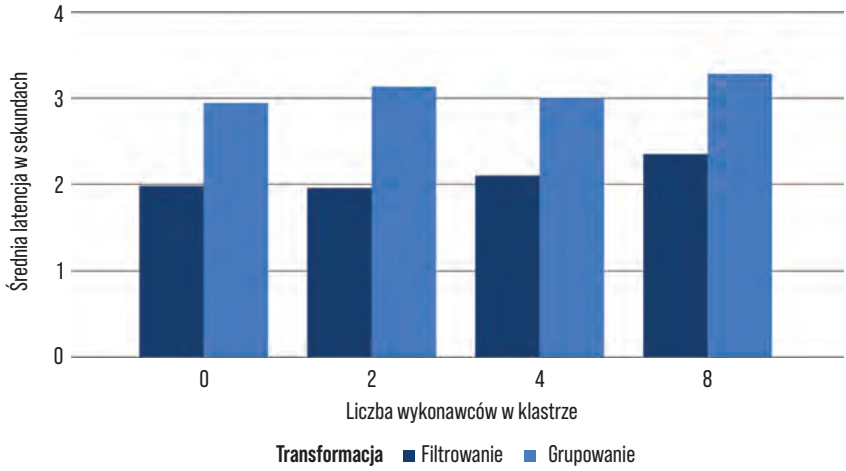
Źródło: opracowanie własne.

Rysunek 9.4 przedstawia średnią latencję dla obu transformacji danych w zależności od liczby wykonawców w klastrze. Dla każdej wielkości klastra średnia latencja jest konsekwentnie niższa przy wykonywaniu filtrowania w porównaniu z grupowaniem. Jest to związane z dwoma czynnikami. Po pierwsze, grupowanie jest szeroką transformacją i wymaga przetasowania danych. Po drugie, grupowanie jest również transformacją z pamięcią stanu (*stateful transformation*), co oznacza, że pomiędzy kolejnymi wsadami danych przechowywany jest ich stan w pamięci. Wykres wskazuje również, że najlepsza średnia latencja występuje dla klastra jednowęzłowego przy grupowaniu oraz dla klastra jednowęzłowego i klastra z dwoma wykonawcami przy filtrowaniu. Dla obu transformacji największy klaster ma najwyższą średnią latencję.

Rysunek 9.4 sugeruje również, że klaster jednowęzłowy może stanowić wystarczającą infrastrukturę dla testowanych scenariuszy. Aby potwierdzić to założenie, potrzebna jest bardziej szczegółowa analiza danych. Dlatego po analizie średniej latencji dla

transformacji danych dla danej wielkości klastra zbadaliśmy średnią latencję dla każdego testu. Poszczególne testy różnią się od siebie na podstawie różnych konfiguracji strumienia, takich jak rozmiar wsadu danych (*batch size*) lub liczba partycji Kafki.

Rysunek 9.4. Średnia latencja dla grupowania i filtrowania



Źródło: opracowanie własne.

Tabela 9.2. Najlepsza i najgorsza latencja dla każdego scenariusza

Transformacja	Prędkość przepływu danych	Najlepsza średnia latencja		Najgorsza średnia latencja	
		liczba wykonawców	średnia latencja w sekundach	liczba wykonawców	średnia latencja w sekundach
Filtrowanie	600	4	1,72	8	10,19
Filtrowanie	1200	2	1,81	4	4,16
Filtrowanie	1800	0	1,79	4	5,05
Filtrowanie	3000	4	1,60	8	5,17
Grupowanie	600	4	2,34	2	4,24
Grupowanie	1200	4	2,66	4	3,51
Grupowanie	1800	0	2,59	4	3,89
Grupowanie	3000	4	2,43	0	3,60

Źródło: opracowanie własne.

Wybór najlepszych i najgorszych wyników dla każdego scenariusza przedstawiono w tabeli 9.2. Średnia latencja dla optymalnej konfiguracji waha się od 1,5 do 2,7 sekundy, a dla najgorszej konfiguracji od 3,5 do nawet 10 sekund. Dla dwóch

z ośmiu scenariuszy (oba przy prędkości przepływu danych 1800) klaster jednowęzłowy wykazuje najlepszą średnią latencję.

Dla pozostałych scenariuszy (dla prędkości 600, 1200 i 3000) przeanalizowaliśmy najlepszą średnią latencję dla pojedynczego testu przeprowadzonego na klastrze jednowęzłowym, aby sprawdzić, jak dokonać jego porównania z klastrem, który uzyskał najlepsze rezultaty. Tabela 9.3 przedstawia porównanie najlepszej średniej latencji dla pojedynczego testu wykonanego na wygrywającym klastrze z najlepszą średnią latencją dla pojedynczego testu przeprowadzonego na klastrze jednowęzłowym – obydwa dla tego samego scenariusza.

Ostatnia kolumna w tabeli 9.3 pokazuje różnicę między średnią latencją najlepszego testu wykonanego na klastrze jednowęzłowym a średnią latencją testu przeprowadzonego na wygrywającym klastrze. Różnica jest nieistotna dla połowy scenariuszy, a dla pozostałych wynosi mniej niż 500 milisekund. W dalszym ciągu jest to zgodne z naszym założeniem, aby unikać nadmiernego rozbudowywania infrastruktury.

Tabela 9.3. Najlepsza średnia latencja dla klastra wygrywającego oraz dla klastra jednowęzłowego

Transformacja	Prędkość przepływu danych	Liczba wykonawców	Najlepsza średnia latencja w sekundach dla wygrywającego klastra	Najlepsza średnia latencja w sekundach dla klastra jednowęzłowego	Różnica w sekundach
Filtrowanie	600	4	1,72	1,96	0,24
Filtrowanie	1200	2	1,81	1,87	0,06
Filtrowanie	1800	0	1,79	nie dotyczy	nie dotyczy
Filtrowanie	3000	4	1,60	1,93	0,33
Grupowanie	600	4	2,34	2,60	0,26
Grupowanie	1200	4	2,66	2,67	0,01
Grupowanie	1800	0	2,59	nie dotyczy	nie dotyczy
Grupowanie	3000	4	2,43	2,48	0,05

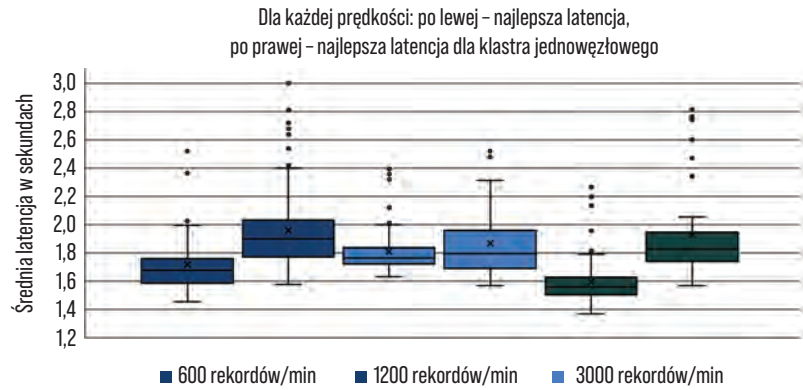
Źródło: opracowanie własne.

W kolejnym kroku poddaliśmy analizie rozkład danych dotyczących latencji dla każdego testu. Rysunki 9.5 i 9.6 pokazują rozkład latencji odpowiednio dla scenariuszy z filtrowaniem i grupowaniem. Każdy rysunek obejmuje tylko te scenariusze, dla których klaster jednowęzłowy nie osiągnął najlepszego rezultatu (scenariusze dla prędkości: 600, 1200 i 3000 rekordów na minutę). Dla każdej prędkości przepływu danych przedstawiono dwa wykresy pudełkowe. Wykres po lewej stronie przedstawia rozkład danych dla testu przeprowadzonego na wygrywającym klastrze, a po prawej –

rozkład latencji dla testu przeprowadzonego na najlepiej skonfigurowanym klastrze jednowęzłowym. Znak „x" oznacza średnią latencję.

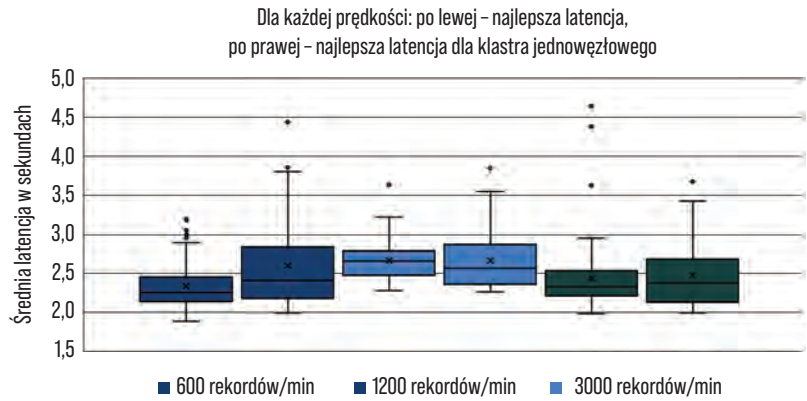
Testy przeprowadzone na klastrze jednowęzłowym wykazują szerszy rozkład danych, co może wskazywać, że infrastruktura jest mniej stabilna. Jednak w przypadku scenariuszy z filtrowaniem 75% danych nadal oscyluje poniżej 2 sekund. Co więcej, w przypadku dwóch scenariuszy z grupowaniem trzeci kwartyl rozkładu jest tylko nieznacznie wyższy od tych dla wygrywających klastrów.

Rysunek 9.5. Rozkład danych najlepszej średniej latencji dla filtrowania – zestawienie najlepszej średniej latencji ogółem z najlepszą średnią latencją dla klastra jednowęzłowego dla danej prędkości



Źródło: opracowanie własne.

Rysunek 9.6. Rozkład danych najlepszej średniej latencji dla grupowania – zestawienie najlepszej średniej latencji ogółem z najlepszą średnią latencją dla klastra jednowęzłowego dla danej prędkości

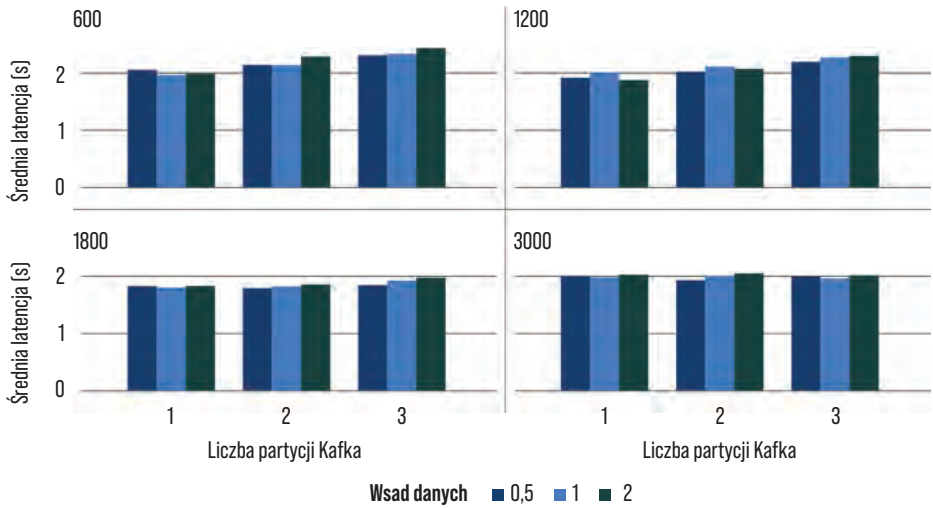


Źródło: opracowanie własne.

Najgorszy wynik odnotowano dla scenariusza z grupowaniem i najniższą prędkością przepływu danych: trzeci kwartyl latencji zmienia się z 2,45 do 2,84 sekundy. W przypadku, gdy stała wydajność nie jest priorytetem, wyniki klastra jednowęzłowego mogą być zadowalające, a wybór mniejszego klastra może okazać się opłacalną strategią.

Rysunki 9.7 i 9.8 przedstawiają po cztery wykresy odpowiednio dla transformacji filtrowania i grupowania. Każdy z nich odpowiada innej prędkości przepływu danych. Z kolei każdy słupek przedstawia średnią latencję w zależności od wielkości wsadu danych po stronie konsumenta strumienia danych i liczby partycji Kafki. Wielkość wsadu danych (0,5 s, 1 s lub 2 s) wyzwala mikro-wsad w regularnych odstępach czasowych i następuje przetwarzanie przyjętych danych we wsadzie. Ponadto po stronie producenta strumienia danych ustawiliśmy liczbę partycji Kafki.

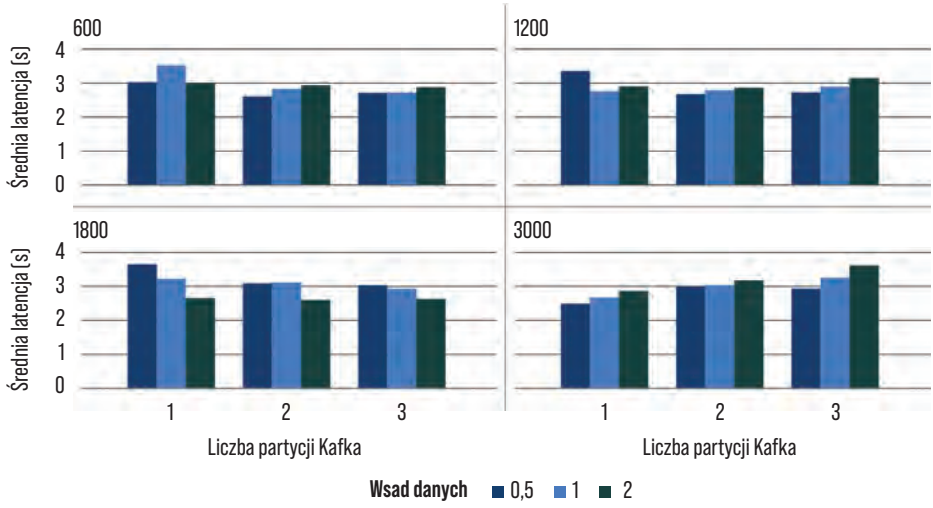
Rysunek 9.7. Średnia latencja dla filtrowania dla klastra jednowęzłowego



Źródło: opracowanie własne.

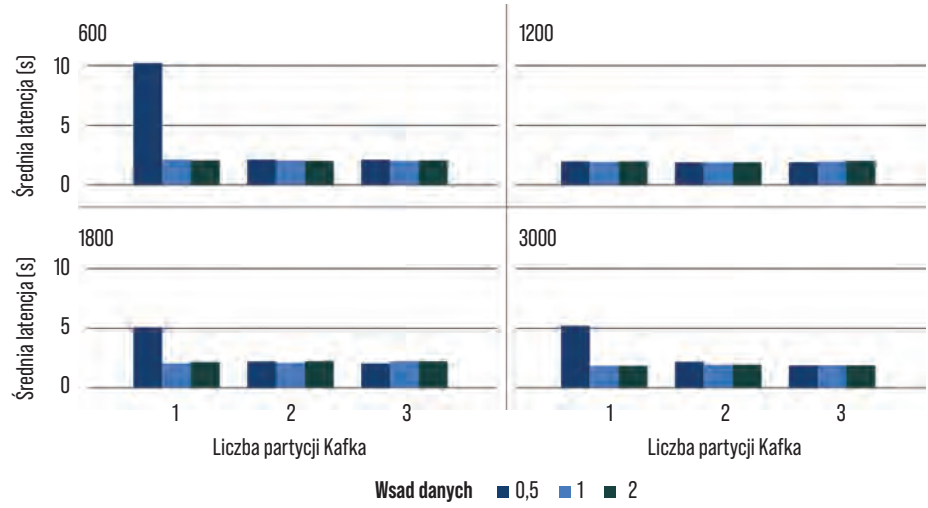
Rysunek 9.7 wskazuje, że w przypadku filtrowania latencja jest najniższa, gdy strumienie odczytują dane z tematu Kafki z jedną partycją. Tendencja ta stopniowo maleje przy wyższych prędkościach przepływu danych. Przy prędkości 3000 rekordów na minutę nie zaobserwowano znaczących zmian w latencji w różnych konfiguracjach. Natomiast w przypadku scenariuszy z grupowaniem zmiany latencji nie są tak spójne pomiędzy różnymi prędkościami danych. Najwyraźniejszy trend jest widoczny przy najwyższej prędkości. Zmniejszenie liczby partycji Kafki i zmniejszenie rozmiaru wsadu danych po stronie konsumenta prowadzi do spadku latencji. Pojedyncza zmiana ustawień może zmniejszyć latencję nawet o 25%.

Rysunek 9.8. Średnia latencja dla grupowania dla klastra jednowęzłowego



Źródło: opracowanie własne.

Rysunek 9.9. Średnia latencja dla filtrowania dla klastra z 8



Źródło: opracowanie własne.

Dodatkowo rysunek 9.9 przedstawia interesującą obserwację dla klastra z ośmioma wykonawcami w scenariuszach z filtrowaniem. Wskazuje on, że średnia latencja dla strumieni danych o wielkości wsadu danych co 500 milisekund oraz przy najniższej prędkości danych gwałtownie wzrasta do 10 sekund. Odpowiednia mediana laten-

cji wynosi 4 sekundy, ale wciąż jest dwa razy wyższa niż mediana latencji dla innych rozmiarów wsadu przy tej samej prędkości danych. Jest to najwyższa średnia latencja zaobserwowana w badaniu. W przypadku rozdzielania mniejszej liczby rekordów strumieniowych na większą liczbę węzłów może pojawić się znaczący narzut operacyjny.

Podsumowanie

W niniejszym artykule opisano przetestowanie różnych scenariuszy przetwarzania danych przy użyciu różnych ustawień konfiguracyjnych. Przede wszystkim wyniki wskazują, że zmniejszenie liczby partycji przetasowania danych podczas grupowania może znacznie poprawić wydajność zapytań. Ponadto transformacje bez pamięci stanu okazały się mniej czasochłonne niż transformacje z pamięcią stanu. Średnio klaster z ośmioma wykonawcami wypadł najgorzej. W poszczególnych testach klaster jednowęzłowy osiągnął najniższą średnią latencję dla niektórych scenariuszy przy optymalnej konfiguracji. Dla innych scenariuszy wykazano, że średnia latencja jest wyższa maksymalnie o 330 milisekund od latencji dla wygrywającego klastra. Te analizy wskazują, że wybór lżejszej infrastruktury może być opłacalny, nawet jeśli wiąże się to z większymi wahaniami latencji. Biorąc pod uwagę wyniki dla klastra jednowęzłowego, zaleca się również zbadanie różnych wartości zarówno dla wielkości wsadu danych, jak i liczby partycji Kafki. Czynniki te mogą wpływać na latencję.

Wyniki badania mają na celu pomóc projektantom architektury w podejmowaniu optymalnych decyzji. Dostarczają wskazówek dotyczących wyboru odpowiedniej wielkości klastra Apache Spark i ustawień parametrów. Co więcej, zapewniają wgląd w możliwe skutki różnych scenariuszy strumieniowania. Optymalizując alokację zasobów i zadania przetwarzania w Apache Spark, organizacje mogą znacząco zredukować zużycie zasobów obliczeniowych oraz czas potrzebny na wnioskowanie, co prowadzi do obniżenia kosztów operacyjnych. Ponadto Spark dostrojony pod kątem wydajności zapewnia elastyczność w tworzeniu i wdrażaniu modeli AI, wykorzystujących dane w czasie rzeczywistym.

Rozważając wyniki tego badania, należy wziąć pod uwagę pewne ograniczenia. Po pierwsze, w naszym badaniu uwzględniliśmy dwa rodzaje transformacji: grupowanie i filtrowanie. Inne transformacje, takie jak liczenie unikalnych wartości czy wykrywanie anomalii, z pewnością byłyby wartościowe. Po drugie, skupiliśmy się wyłącznie na technologiach Apache Spark i Apache Kafka. Ciekawe byłoby porównanie tych technologii z innymi technologiami (np. systemami opartymi na chmurze).

Kierunki dalszych badań wynikają z ograniczeń tego badania. Uwzględnienie bardziej złożonych transformacji i analiz lub rozszerzenie zakresu analizowanych technologii z pewnością zwiększyłoby praktyczną wartość badania.

Bibliografia

Ahmed, N., Barczak, A.L.C., Rashid, M.A., Susnjak, T. (2021). An Enhanced Parallelisation Model for Performance Prediction of Apache Spark on a Multinode Hadoop Cluster, *Big Data and Cognitive Computing*, 5(4). DOI: 10.3390/bdcc5040065.

Amazon (2019). *Amazon EMR*, https://aws.amazon.com/emr/ (dostęp: 18.10.2019).

Apache Spark (2020a). *Spark Configuration*, https://spark.apache.org/docs/latest/configuration.html (dostęp: 14.05.2021).

Apache Spark (2020b). *Unified Engine for Large-Scale Data Analytics*, https://spark.apache.org/ (dostęp: 22.07.2021).

Chambers, B., Zaharia, M. (2018). *Spark: The Definitive Guide Big Data Processing Made Simple*, https://learning.oreilly.com/library/view/spark-the-definitive/9781491912201/ (dostęp: 10.09.2022).

Chao, Z., Shi, S., Gao, H., Luo, J., Wang, H. (2018). A Gray-Box Performance Model for Apache Spark, *Future Generation Computer Systems*, *89*, s. 58–67. DOI: 10.1016/j.future.2018.06.032.

Cheng, G., Ying, S., Wang, B. (2021). Tuning Configuration of Apache Spark on Public clouds by Combining Multi-Objective Optimization and Performance Prediction Model, *Journal of Systems and Software*, *180*, 111028. DOI: 10.1016/j.jss.2021.111028.

Damji, J.S., Wenig, B., Das, T., Lee, D. (2020). Learning Spark: Lightning-Fast Data Analytics. W: *Learning Spark: Lightning-Fast Data Analytics* (2nd ed.). Sebastopol: O'Reilly Media.

Databricks (2023). *Apache Spark™*, https://www.databricks.com/spark/about (dostęp: 23.10.2023).

Google Cloud (2022). *Dataproc*, https://cloud.google.com/dataproc (dostęp: 12.10.2022).

IBM (2021). *TeraSort Benchmark*, https://www.ibm.com/docs/en/spectrum-symphony/7.2.1?topic=mapreduce-terasort-benchmark (dostęp: 22.10.2022).

Karau, H., Warren, R. (2017). *High Performance Spark*. Sebastopol: O'Reilly Media.

Mavridis, I., Karatza, H. (2017). Performance Evaluation of Cloud-Based Log File Analysis with Apache Hadoop and Apache Spark, *Journal of Systems and Software*, 125, s. 133–151. DOI: 10.1016/j.jss.2016.11.037.

Petridis, P., Gounaris, A., Torres, J. (2017). Spark Parameter Tuning via Trial-and-Error, *Advances in Intelligent Systems and Computing*, 529, s. 226–237. DOI: 10.1007/978-3-319-47898.

Petrov, M., Butakov, N., Nasonov, D., Melnik, M. (2018). Adaptive Performance Model for Dynamic Scaling Apache Spark Streaming, *Procedia Computer Science*, 136, s. 109–117. DOI: 10.1016/j.procs.2018.08.243.

Salloum, S., Dautov, R., Chen, X., Peng, P.X., Huang, J.Z. (2016). Big Data Analytics on Apache Spark, *International Journal of Data Science and Analytics*, 1(3–4), s. 145–164. DOI: 10.1007/s41060-016-0027-9.

Samadi, Y., Zbakh, M., Tadonki, C. (2018). Performance Comparison Between Hadoop and Spark Frameworks Using HiBench Benchmarks, *Concurrency and Computation: Practice and Experience*, 30(12). DOI: 10.1002/cpe.4367.

Wang, K., Khan, M.M.H. (2015). Performance Prediction for Apache Spark Platform. W: *Proceedings – 2015 IEEE 17th International Conference on High Performance Computing and Communications, 2015 IEEE 7th International Symposium on Cyberspace Safety and Security and 2015 IEEE 12th International Conference on Embedded Software and Systems* (s. 166–173). DOI: 10.1109/HPCC-CSS-ICESS.2015.246.

10.0

KWANTOWE ALGORYTMY HYBRYDOWE JAKO MODELE UCZENIA MASZYNOWEGO

Sebastian Zając
Szkoła Główna Handlowa w Warszawie

Jacob Cybulski
Deakin University Melbourne

Tomasz Kulpa
Uniwersytet Kardynała Stefana Wyszyńskiego w Warszawie

Streszczenie

W rozdziale zaprezentowano kwantowe algorytmy hybrydowe. Pozwalają one realizować klasyczne modele uczenia maszynowego z wykorzystaniem obwodów kwantowych. Po wprowadzeniu podstawowych pojęć przedstawiono algorytm kwantowej sieci neuronowej, który można wykorzystać zarówno do procesu predykcji wartości ciągłej, jak i w procesie klasyfikacji.

Słowa kluczowe: obliczenia kwantowe, kwantowe uczenie maszynowe, sztuczna inteligencja

Wprowadzenie

Wraz z dynamicznym wzrostem ilości generowanych danych rośnie zapotrzebowanie na moc obliczeniową, umożliwiającą przetwarzanie tych danych w skończonym czasie, oraz wspieranie podejmowania decyzji biznesowych. Pomimo postępu technologicznego, w tym dostępu do coraz tańszego i bardziej wydajnego sprzętu, wiele problemów obliczeniowych wciąż pozostaje nierozwiązywalnych w rozsądnym czasie. Tradycyjnie wyzwania te dotyczyły głównie fizyki i chemii, jednak współczesne zastosowania biznesowe, takie jak optymalizacja i sztuczna inteligencja, również wymagają narzędzi o ogromnej mocy obliczeniowej. Obecne komputery często nie są w stanie sprostać tym wymaganiom. Technologią, która próbuje rozwiązać problemy obliczeniowe klasycznych komputerów, są komputery kwantowe, stosowane w takich dziedzinach, jak kryptografia, symulacje fizyczne i chemiczne, uczenie maszynowe i optymalizacja. Komputery kwantowe nie są jeszcze maszynami, które moglibyśmy wykorzystać do codziennych zadań. Te dostępne pozwalają wykonywać algorytmy i obliczenia na setkach kubitów. Nie posiadają jednak pełnego mechanizmu korekcji błędów. Ponadto bramki muszą działać znacznie szybciej niż ich czas dekoherencji, co uniemożliwia realizację długich sekwencji bramek dla złożonych algorytmów. Dlatego obecny etap rozwoju tych maszyn nazywany jest erą **NISQ** (*noisy intermediate-scale quantum)* [Preskill, 2018]. Pomimo ograniczeń technologicznych wciąż można wykazać tzw. kwantową supremację (czyli przewagę algorytmów kwantowych nad klasycznymi) w problemach optymalizacyjnych i w modelowaniu danych, wykorzystując do tego celu parametryzowane obwody kwantowe (*parameterized quantum circuits*, PQCs), które z zastosowaniem klasycznych optymalizatorów mogą być trenowane w celu znalezienia optymalnych wartości dla zadanej funkcji kosztu. Podejście takie nazywane jest uczeniem hybrydowym. PQC realizowane są z użyciem bramek w postaci ustalonej (np. bramki CNOT). Wykorzystują one również bramki parametryzowane, co pozwala generować nietrywialne wyniki [Lund, 2017; Harrow, 2017]. Modele kwantowego uczenia maszynowego (*quantum machine learning*, QML) realizowane przez kwantowe algorytmy wariacyjne (*variational quantum algorithms*, VQA) reprezentują całą klasę algorytmów, które używają klasycznych optymalizatorów do znalezienia parametrów kwantowych obwodów. Szczególnymi realizacjami tak zdefiniowanych modeli są: *variational quantum eigensolver* [Peruzzo, 2014], *variational quantum solvers* [Cerezo, 2021], *variational quantum classifier* [Havlicek, 2019], *quantum support vector classification* [Hsu, 2020], *quantum neural networks* [Benedetti, 2019], *quantum autoencoder* [Cybulski, 2024] czy też *quantum approximate optimization algorithm*, stosowane w zadaniach optymalizacyjnych typu QUBO [Farhi, 2014].

Wszystkie tego typu algorytmy można realizować w bibliotekach Pythona, takich jak IBM Qiskit, Pennylane, Cirq z wykorzystaniem prawdziwych komputerów kwantowych, udostępnianych przez dostawców chmurowych, typu AWS, Google Quantum AI, Azure Quantum, IBM Q Experience, Leap (D-Wave) i wielu innych.

10.1. Modele obliczeń kwantowych

Model obliczeń kwantowych znacząco różni się od klasycznego podejścia. Aby zrozumieć te różnice, wprowadźmy podstawowe pojęcia i terminologię stosowaną zarówno w obliczeniach klasycznych, jak i kwantowych.

Klasyczne komputery przetwarzają informacje zakodowane w postaci ciągu bitów. Każdy bit przyjmuje jedną z dwu wartości: 0 lub 1. Tranzystory, będące podstawowym elementem współczesnych komputerów, umożliwiają zarówno zapis informacji (utworzenie stanu początkowego), jak i jej odczyt (weryfikację stanu końcowego). Proces obliczeń realizowany jest poprzez zastosowanie bramek logicznych, które przekształcają stan początkowy bitów w stan końcowy zgodnie z zaprogramowanymi regułami.

Teoretyczny opis działania klasycznych komputerów opiera się na koncepcjach zaproponowanych przez Turinga [1950] oraz von Neumanna [1945]. Alan Turing wprowadził pojęcie maszyny Turinga jako abstrakcyjnego modelu obliczeniowego, który może wykonać dowolny algorytm, jeśli zostanie odpowiednio zaprogramowany. Natomiast John von Neumann opisał architekturę współczesnych komputerów, opartą na centralnym procesorze (*central processing unit*, CPU), pamięci oraz jednostkach wejścia – wyjścia. Szczegółowe informacje na temat działania komputerów klasycznych można znaleźć w literaturze [Wong, 2022; Feynman, 2022]. Publikacje te dostarczają wiedzy zarówno teoretycznej, jak i praktycznej na temat klasycznych modeli obliczeniowych.

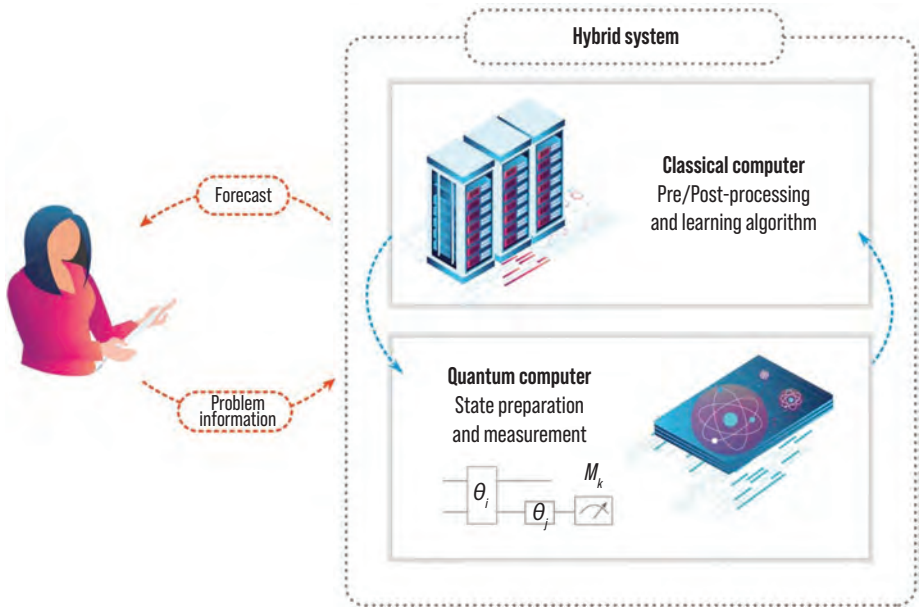
W obszarze obliczeń kwantowych wyróżnia się trzy podstawowe modele, które różnią się metodami realizacji obliczeń. Są to:

- **obwody kwantowe** (*quantum circuits*), wykorzystujące bramki kwantowe do wykonywania operacji na kubitach [Nielsen, Chuang, 2010];
- **adiabatyczne obliczenia kwantowe,** zaproponowane m.in. przez firmę D-Wave, pozwalające rozwiązywać problemy optymalizacyjne z zastosowaniem kwantowego wyżarzania [Hauke *et al.*, 2020];
- **topologiczne komputery kwantowe** rozwijane m.in. przez firmę Microsoft, bazujące na zaawansowanej topologii algebraicznej oraz oferujące innowacyjne podejście do realizacji obliczeń kwantowych [Nayak *et al.*, 2008].

W niniejszej pracy skoncentrujemy się wyłącznie na modelu opartym na obwodach kwantowych.

Pierwszy procesor kwantowy został opracowany w 1996 r. [Gershenfeld, 1996]. Urządzenie to, bazujące na technologii magnetycznego rezonansu jądrowego (*nuclear magnetic resonance*, NMR), wykorzystywało impulsy radiowe do realizacji bramek kwantowych i obsługiwało dwa kubity. W 2016 r. firma IBM wprowadziła 5-kubitowy procesor kwantowy dostępny w środowisku chmurowym. Obecnie, w 2025 r., dostępne są komputery kwantowe umożliwiające wykonanie obliczeń na ponad 1000 kubitach.

Rysunek 10.1. Schemat hybrydowego procesu obliczeń kwantowych



Źródło: Benedetti [2019].

Proces obliczeń kwantowych, zilustrowany na rysunku 10.1, wymaga współpracy klasycznego komputera z procesorem kwantowym. Klasyczny komputer inicjalizuje stan początkowy kubitów za pomocą instrukcji programistycznych, definiuje konfigurację bramek kwantowych, które realizują dany algorytm, oraz przesyła te ustawienia do procesora kwantowego. Procesor kwantowy przetwarza algorytm, modyfikując stan kubitów zgodnie z programem. Po zakończeniu przetwarzania kubity są mierzone, a wyniki pomiarów przesyłane z powrotem do klasycznego komputera, który rejestruje i analizuje dane. Proces ten można porównać do działania procesorów graficznych (*graphics processing unit*, GPU) wykorzystywanych w zadaniach, takich jak wyznacza-

nie parametrów sieci neuronowych. Dane wejściowe, przedstawione w postaci macierzy, są ładowane do karty graficznej jako bity, gdzie GPU realizuje obliczenia poprzez operacje logiczne. Po zakończeniu obliczeń wyniki są odczytywane i interpretowane przez komputer wyposażony w jednostkę centralną przetwarzania.

10.2. Definicje matematyczne wykorzystywane w obliczeniach kwantowych

Kubit (*quantum bit*) jest podstawowym nośnikiem informacji w obliczeniach kwantowych. Fizycznie może on być realizowany jako dwustanowy układ kwantowy. Stan kubitu opisujemy jako wektor w przestrzeni Hilberta. Opis ten może dotyczyć zarówno pojedynczego kubitu, jak i całego zbioru kubitów. Aby wyjaśnić kluczowe koncepcje algebry liniowej stosowanej w obliczeniach kwantowych, wykorzystujemy notację Diraca. W tej notacji wektory są reprezentowane przez symbole: bra „$\langle \cdot |$" oraz ket „$| \cdot \rangle$".

Stan kubitu $|\psi\rangle$ jest opisywany za pomocą superpozycji stanów $|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ i $|1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ jako wektor w postaci kolumnowej. W przypadku dwuwymiarowym wektor „ket" można zapisać jako:

$$|\psi\rangle = \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} = \alpha_0|0\rangle + \alpha_1|1\rangle, \tag{10.1}$$

gdzie α_0, $\alpha_1 \in C$ są liczbami zespolonymi.

Wartości α_0, α_1 spełniają własność unormowania

$$|\alpha_0|^2 + |\alpha_1|^2 = 1. \tag{10.2}$$

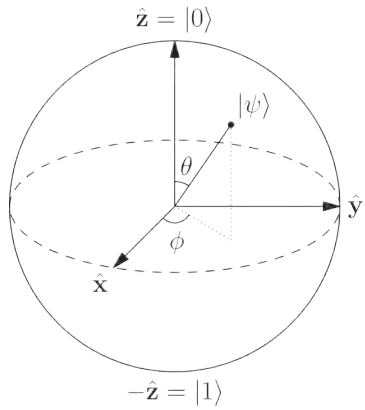
Zespolone współczynniki α_0, α_1 nazywamy **amplitudami**. Wykorzystując unormowanie 10.2 oraz fakt, iż istotne są tylko kwadraty amplitud (interpretowane jako prawdopodobieństwo), można przedstawić $|\psi\rangle$ jako:

$$|\psi\rangle = \cos\frac{\theta}{2}|0\rangle + e^{i\phi}\sin\frac{\theta}{2}|1\rangle, \tag{10.3}$$

gdzie θ, $\phi \in R$.

Pozwala to zaprezentować stan kubitu jako punkt na sferze Blocha.

Rysunek 10.2. Sfera Blocha – graficzna reprezentacja stanu pojedynczego kubitu



Źródło: opracowanie własne.

Wektory„bra" tworzone są jako sprzężenie hermitowskie i transpozycja wektorów ket.

$$\langle\psi| = \alpha_0^*\langle 0| + \alpha_1^*\langle 1| \;;\; \langle\psi| = |\psi\rangle^\dagger = [\alpha_0^* \;\; \alpha_1^*] \tag{10.4}$$

Notacja Diraca pozwala na „wygodny zapis" działań na wektorach stanu, takich jak **iloczyny skalarne** $\langle\psi|\phi\rangle$ lub **iloczyny tensorowe** służące do opisu systemów wielo-kubitowych.

Dla dwóch różnych kubitów: $|\psi\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$ oraz $|\phi\rangle = \beta_0|0\rangle + \beta_1|1\rangle$ iloczyn skalarny zdefiniowany jest jako:

$$\langle\psi|\phi\rangle = \alpha_0^*\beta_0\langle 0|0\rangle + \alpha_0^*\beta_1\langle 0|1\rangle + \alpha_1^*\beta_0\langle 1|0\rangle + \alpha_1^*\beta_1\langle 1|1\rangle = \alpha_0^*\beta_0 + \alpha_1^*\beta_1 \tag{10.5}$$

Iloczyn tensorowy wyraża się przez:

$$|\psi\rangle \otimes |\phi\rangle = \alpha_0\beta_0|00\rangle + \alpha_0\beta_1|01\rangle + \alpha_1\beta_0|10\rangle + \alpha_1\beta_1|11\rangle, \tag{10.6}$$

gdzie $\sum_{i,j=0}^{1} |\alpha_i\beta_j|^2 = 1$.

W przypadku n klasycznych bitów istnieje 2^n możliwych różnych stanów klasycznych:

$$00..0, 00..1, \ldots, 1..11. \;\; (0, 1, \ldots, 2^n - 1) \tag{10.7}$$

Posiadając n kubitów, możemy utworzyć stan realizowany przez 2^n wektorów bazowych i tyle samo amplitud:

$$|00..0\rangle, |00..1\rangle, \ldots, |1..11\rangle, \ (|0\rangle, |1\rangle, \ldots, |2^n - 1\rangle) \tag{10.8}$$

n kubitów opisuje stan naszego komputera kwantowego jako wektor jednostkowy, należący do przestrzeni generowanej iloczynem tensorowym (10.6) $\mathbb{C}^{2^n}$. Obliczenia realizowane są zazwyczaj od przygotowania stanu podstawowego, wyrażanego jako $|00..0\rangle = |0\rangle^{\otimes n}$. Dla uproszczenia zapisu tensorowego można ten stan określić jako $|0\rangle$.

Często można spotkać się z opinią, że **superpozycja** stanowi główną przewagę komputerów kwantowych nad klasycznymi. W rzeczywistości samo istnienie superpozycji nie wystarcza do rozwiązania wszystkich problemów obliczeniowych. Kluczowym wyzwaniem w praktycznym wykorzystaniu superpozycji jest proces pomiaru. W trakcie pomiaru stan kubitu nie jest bezpośrednio obserwowany w postaci superpozycji, lecz przechodzi w jeden ze stanów bazowych. Dla kubitu opisanego równaniem 10.8 wynikiem pomiaru może być stan $|0\rangle$ z prawdopodobieństwem $|\alpha_0|^2$ lub stan $|1\rangle$ z prawdopodobieństwem $|\alpha_1|^2$. W fizyce nazywamy ten efekt kolapsem wektora stanu.

W kontekście przetwarzania informacji mamy układ kwantowy, którego stan potrafimy przygotować i zmierzyć. Zostaje nam zatem opis operacji na kubitach. Do tego celu wprowadzimy pojęcie **operatora liniowego**. W ogólnym ujęciu operator A działa na stan i przekształca go w jakiś inny stan (z tej samej przestrzeni).

$$A|\psi\rangle = |\phi\rangle \tag{10.9}$$

W obliczeniach kwantowych będzie interesować nas szczególny rodzaj operatorów zdefiniowanych przez równanie własne:

$$A|\psi_a\rangle = a|\psi_a\rangle, \tag{10.10}$$

gdzie $|\psi_a\rangle$ nazywamy stanem własnym operatora A z wartością własną $a \in R$.

Operator A reprezentuje działanie bramki kwantowej. Ponieważ każdy kubit jest dwuwymiarowym wektorem, bramki (działające na n-kubitów) reprezentowane będą przez macierze unitarne $M_{2^n}(\mathbb{C})$ o wartościach zespolonych. Własność unitarności $U^\dagger U = UU^\dagger = I$ pozwala zachowywać prawdopodobieństwo (kwadrat normy wektora stanu). Unitarność operatorów w obliczeniach kwantowych wprowadza dodatkową, istotną cechę bramek – ich **odwracalność**. Przykładem nieparametryzowanych bramek działających na jeden kubit mogą być macierze Pauliego czy też bramka Hadamarda:

$$\sigma_x = X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},\ \sigma_y = Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix},$$
$$\sigma_z = Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},\ H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{10.11}$$

Tak zdefiniowany operator można zapisać w postaci $A = a_i |\psi_i\rangle\langle\psi_i|$, gdzie a_i to wartości własne, natomiast $|\psi_i\rangle\langle\psi_i|$ to operatory rzutowe odpowiadające przestrzeni i-tego wektora własnego. Wynik pomiaru stanu komputera kwantowego odpowiada jednej z wartości własnej. Obliczając:

$$\langle\psi|A|\psi\rangle = \sum_{i,j} \langle\psi|\psi_i\rangle\langle\psi_i|\psi\rangle\langle\psi_i|A|\psi_i\rangle = \sum_i |\langle\psi|\psi_i\rangle|^2 a_i = \sum_i p(\psi_i) a_i \tag{10.12}$$

otrzymujemy wartość oczekiwaną operatora A w stanie $|\psi\rangle$, jeśli tylko $|\langle\psi|\psi_i\rangle|^2 = p(\psi_i)$ określimy jako prawdopodobieństwo, że po pomiarze nasz układ będzie w stanie własnym $|\psi_i\rangle$.

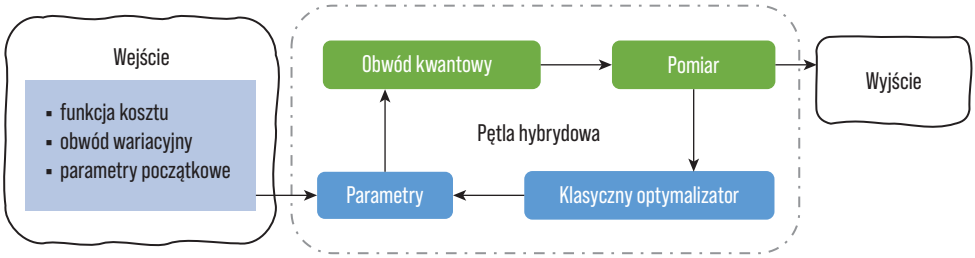
Posiadając zestaw bramek działających na kubity, możemy je wykorzystać do budowy **obwodów kwantowych**, stanowiących fundament realizacji algorytmów kwantowych. Te obwody składają się z ciągu operacji, które przekształcają stan początkowy kubitów za pomocą odwracalnych bramek kwantowych. Wykorzystanie takich obwodów może tworzyć przewagę nad klasycznymi algorytmami np. algorytm Deutsch-Josza [1992], Grovera [1996] czy Shora [1997].

Bramki mogą być zależne od parametrów, co pozwala nam tworzyć parametryzowane układy kwantowe. Przykładowe bramki przedstawia równanie 10.13.

$$R_X(\theta) = \begin{bmatrix} \cos\frac{\theta}{2} & -i\sin\frac{\theta}{2} \\ -i\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix},\ R_Y(\theta) = \begin{bmatrix} \cos\frac{\theta}{2} & -\sin\frac{\theta}{2} \\ \sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix},$$
$$R_Z(\theta) = \begin{bmatrix} e^{-i\frac{\theta}{2}} & 0 \\ 0 & e^{i\frac{\theta}{2}} \end{bmatrix} \tag{10.13}$$

Posiadając powyższe elementy, możemy zrealizować model obwodów kwantowych z wykorzystaniem PQC, zwanych też obwodami wariacyjnymi. Pozwala to realizować VQA w celu przybliżonego znajdowania minimum funkcji kosztu. Schemat wariacyjnego algorytmu kwantowego został przedstawiony na rysunku 10.3.

Rysunek 10.3. Schemat wariacyjnego algorytmu kwantowego



Źródło: opracowanie własne.

10.3. Kwantowe uczenie maszynowe a sztuczna inteligencja

Postęp technologiczny i powszechna cyfryzacja sprawiły, że dane stały się powszechnym i cennym zasobem [Zając, 2022]. Są one generowane i przetwarzane zarówno w sposób ustrukturyzowany, jak i nieustrukturyzowany. Strukturyzacja danych doprowadziła do rozwoju wielu modeli, które dziś ogólnie określamy jako modele uczenia maszynowego (*machine learning*, ML). Natomiast przetwarzanie danych nieustrukturyzowanych, takich jak tekst, obrazy czy wideo, przyczyniło się do rozwoju uczenia głębokiego (*deep learning*, DL). Oba te podejścia, często określane zbiorczo jako sztuczna inteligencja (*artificial inteligence*, AI), zostały stworzone głównie do rozpoznawania wzorców. Jednak coraz częściej wykorzystywane są również do modelowania i generowania nowych danych.

Klasyczny model sztucznej inteligencji możemy wyrazić jako funkcję $f(x;\theta)$, która zależy zarówno od danych reprezentowanych przez ustrukturyzowaną macierz x, jak i od parametrów θ, których wartości zostają ustalone w procesie uczenia.

W przypadku modeli kwantowych model uczenia maszynowego realizowany jest jako trzyczęściowy proces:

1. *Preprocessing* – pozwala załadować wektor danych $\vec{x}$ do parametrycznego obwodu kwantowego z wykorzystaniem tzw. *feature* mapy, czyli funkcji $\phi(\vec{x})$. Funkcja ta wyrażana jest jako parametryzowany obwód kwantowy, opisany przez operator $U_{\phi(\vec{x})}|0\rangle$. W tym miejscu warto zwrócić uwagę, iż procedura ta jest podobna do stosowania zanurzeń danych nieustrukturyzowanych w sieciach neuronowych (*embedding*), np. przetworzenie danych tekstowych z wykorzystaniem algorytmu Word2Vec. W fazie tej bardzo ważnym krokiem jest klasyczne przygotowanie danych poprzez takie elementy, jak czyszczenie braków danych, standaryzacja czy też transformacje pozwalające otrzymać odpowiedni rozkład zmiennej.

2. Parametryzowany kwantowy obwód wariacyjny – reprezentowany jest przez operator W_θ zależny od wektora parametrów θ, który działa na wektor stanu przygotowany przez obwód kodujący dane $W_\theta U_{\phi(\vec{x})}|0\rangle$. Ze względu na bardzo dużą ilość kombinacji bramek, które można łączyć równolegle i szeregowo, trudno jest wskazać jeden konkretny obwód pozwalający realizować wszystkie problemy analityczne (nie ma darmowych obiadów). Wybór odbywa się najczęściej poprzez ustalenie różnych schematów i ich trenowania [Sim, 2019]. Porównać można to do klasycznych sieci neuronowych, gdzie ilość warstw ukrytych, liczba neuronów w każdej warstwie czy funkcje aktywacji są tzw. hiper-parametrami, które ustala się przez doświadczenie. Wybrany schemat obwodów kwantowych określa się często mianem ansatzu.
3. Pomiar zrealizowanego obwodu – na tym etapie najczęściej estymuje się zbiór wartości oczekiwanych $\left\{\langle A_k\rangle_{x,\theta}\right\}_{k=1}^{K}$, przyjmujący wartości od −1 do 1. W zależności od realizowanego procesu można wykonać pomiar jednego lub większej ilości kubitów. Można również generować wynik w postaci amplitud, prawdopodobieństw czy też w postaci binarnej. Etap ten często nazywany jest postprocesingiem danych.

Do zakodowania informacji klasyczne komputery używają bitów (które mogą przyjąć wartość 0 lub 1). Aktualnie używa się systemów, które kodują znaki w postaci 32- lub 64-bitowych sekwencji. Pozwala to zakodować wszystkie znaki z kodowania ASCII bądź Unicode. Istnieje wiele różnych sposobów kodowania (*encode*) lub zanurzenia (*embed*) informacji w układ *n*-kubitów opisany przez stan kwantowy. Ponieważ chcemy użyć kwantowego komputera do uczenia się algorytmu na podstawie klasycznych danych, musimy zdecydować, w jaki sposób będziemy reprezentować wiersz danych, a nawet cały zbiór danych w postaci stanu kwantowego. W tradycyjnym ujęciu obliczeń kwantowych proces przygotowania komputera w stan początkowy nazywany jest przygotowaniem stanu (*state preparation*). W przypadku dużej ilości zmiennych można przeprowadzić redukcję wymiarów, wykorzystując np. mechanizm PCA lub inne metody opisane przez Przanowskiego [2021].

Jedną ze strategii jest tzw. kodowanie bazowe (*basis encoding*), które polega na zamianie wektora danych wyrażonego w postaci bitowej na wektory bazowe kubitów ($0 \to |0\rangle$ i $1 \to |1\rangle$). Chcąc jednak kodować dużo zmiennych i z dużą precyzją poszczególnych wartości, będziemy zmuszeni do wykorzystania bardzo dużej ilości kubitów, co dla obecnych maszyn nie jest zbyt dobrym rozwiązaniem. Kodowanie to, tak samo jak w przypadku bitów, pozwala realizować zarówno liczby całkowite, jak i rzeczywiste (dla z góry określonej precyzji).

Układ *n*-kubitów może być reprezentowany jako superpozycja stanów bazowych. Pozwala to zakodować dane w amplitudach (*amplitude encoding*). Normalizując wektor danych $x = \begin{pmatrix} x_1 \\ \vdots \\ x_{2^k} \end{pmatrix}$, tak by $\sum_k |x_k|^2 = 1$, możemy zakodować wszystkie x_k jako amplitudy. Na przykład $x = (0{,}073,\ -0{,}438,\ 0{,}730,\ 0.000) \leftrightarrow |x\rangle = 0{,}073|00\rangle - 0{,}438|01\rangle + 0{,}730|10\rangle + 0|11\rangle$

W niniejszej pracy zastosujemy inne kodowanie. Ponieważ wykorzystuje ono macierze rotacji $R_X(\theta)$, $R_Y(\theta)$, $R_Z(\theta)$ zdefiniowane w równaniu 10.13, kodowanie to nazywane jest kodowaniem rotacyjnym (*angle encoding*). Przetwarza ono zmienne na iloczyny i sumy funkcji cosinus i sinus. W kodowaniu tym istotne jest, aby podczas procesu skalowania zmiennych wyskalować je do wartości (−1,1).

$$|x\rangle = \bigotimes_{i=1}^{N} \cos(x_i)|0\rangle + \sin(x_i)|1\rangle$$

Informacje o pozostałych sposobach kodowania danych można znaleźć w artykułach Schuld, Bocharova, Svore i Wiebe'a [2021] oraz LaRose'a i Coyle'a [2020].

Analogicznie do uniwersalnego twierdzenia granicznego dla sieci neuronowych, zawsze istnieje obwód kwantowy, który może aproksymować funkcję z dowolnie małym błędem. W praktyce wiele wariacyjnych obwodów kwantowych realizowanych jest z ustaloną strukturą bramek, co – pomimo wykładniczego wzrostu ilości amplitud – pozwala na zredukowanie złożoności modeli poprzez małą liczbę wolnych do trenowania parametrów. Przy strategii ustalania szablonu bramek wykorzystywanych do uczenia modelu bierze się pod uwagę fakt, że dzisiejsze komputery kwantowe wciąć bazują na niedużej liczbie kubitów i można na nich realizować bardzo rzadko połączone bramki jedno- lub dwukubitowe. Najczęściej wykorzystuje się pojedyncze bramki obrotów jako elementy parametryzowane oraz bramki CNOT służące do generowania splątania pomiędzy poszczególnymi kubitami. Zarówno obwody wariacyjne, jak i sieci neuronowe mogą być traktowane jako warstwy połączonych składników kontrolowanych poprzez trenowalne parametry. Prowadzi to często do traktowania obwodów wariacyjnych jako kwantowych modeli sieci neuronowych. Obwody kwantowe realizują unitarne i jednocześnie liniowe bramki, podczas gdy istotną cechą sieci neuronowych jest zastosowanie nieliniowych funkcji aktywacji. Jednym ze sposobów realizacji nieliniowości w obwodach kwantowych jest wykorzystanie mechanizmu pomiaru kwantowego.

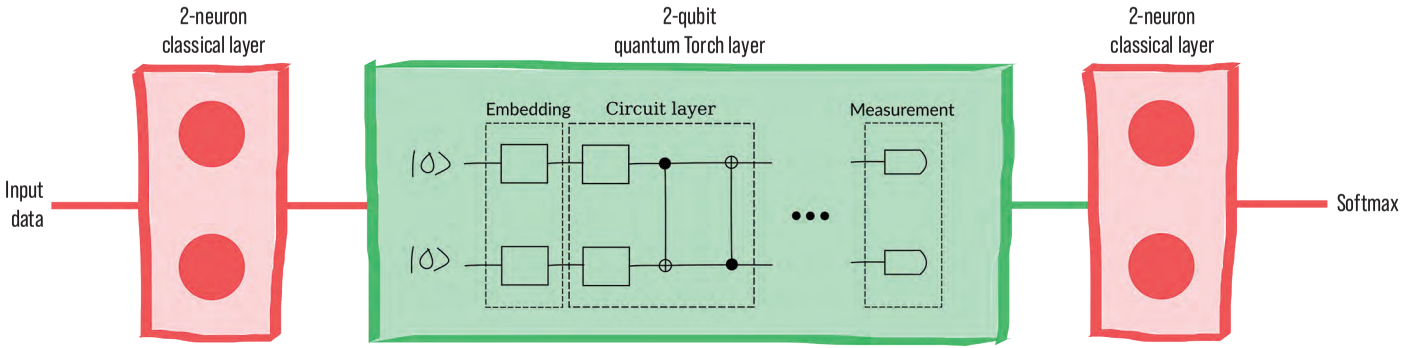
10.4. Hybrydowy algorytm kwantowej sieci neuronowej w procesie klasyfikacji binarnej

Poniżej przedstawiamy realizację procesu klasyfikacji binarnej z wykorzystaniem klasycznej sieci neuronowej oraz modelu hybrydowego, zawierającego klasyczne warstwy wejściowe i wyjściowe dla syntetycznie wygenerowanych danych. Dane wygenerowane zostały przy pomocy funkcji pochodzącej z pakietu scikit-learn *make_moons* (https://shorturl.at/Fjagn) z parametrami *n_sample = 100* oraz *noise = 0,4*. Sieci neuronowe zrealizowano z wykorzystaniem środowiska Python i biblioteki PyTorch. W roli klasycznego optymalizatora zastosowano funkcję Adam z parametrem uczenia 0,01, a jako funkcje straty – binarną entropię krzyżową (*binary cross entropy*). Wszystkie modele uczone były przez 100 epok. Pierwsza sieć neuronowa utworzona została z jedną warstwą liniową (dwie zmienne wejściowe i dwie wyjściowe) wraz z funkcją aktywacji softmax. Druga sieć neuronowa zawiera: warstwę wejściową zakończoną funkcją aktywacji Relu, warstwę ukrytą zakończoną kolejną funkcją aktywacji Relu oraz warstwę wyjściową zakończoną funkcją softmax. Trzecia sieć to hybrydowe połączenie klasycznej liniowej warstwy wejściowej z kwantowym obwodem dwu-kubitowym, zrealizowanym zgodnie ze schematem przedstawionym na rysunku 10.4. Schemat warstwy kwantowej pokazany został na rysunku 10.5.

Schemat czwartej sieci neuronowej, ukazującej możliwość składania warstw kwantowych w sposób równoległy, został zaprezentowany na rysunku 10.6.

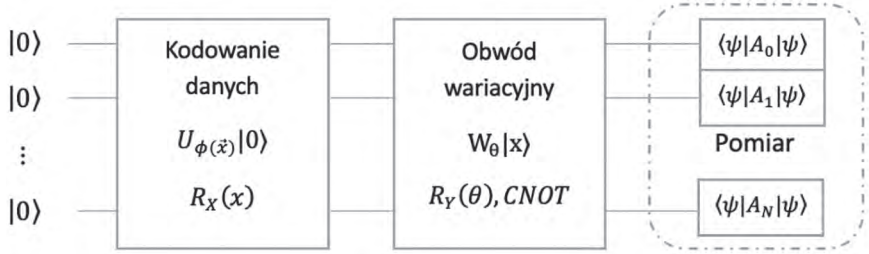
Jakość modeli przedstawiona została na podstawie metryki jakości dopasowania wyników (*accuracy*). Model klasycznej sieci bez warstw ukrytych (realizujący proces regresji logistycznej) osiągnął wynik 83% na zbiorze testowym. Sieć z warstwami ukrytymi wykazała typowy problem przeszacowania (*overfitting*), idealnie dopasowując się do danych treningowych. Jednak na zbiorze testowym osiągnęła zdolność tylko 73%. Pierwsza sieć kwantowa osiagnęła wynik 71% na zbiorze testowym i 84% na zbiorze treningowym. Sieć z warstwami kwantowymi złożonymi równolegle uzyskała 73% na zbiorze testowym i 91% na zbiorze treningowym.

Rysunek 10.4. Schemat hybrydowej kwantowej sieci neuronowej z klasyczną warstwą wejściową, warstwą kwantową zakończoną warstwą wyjściową z funkcją aktywacji softmax



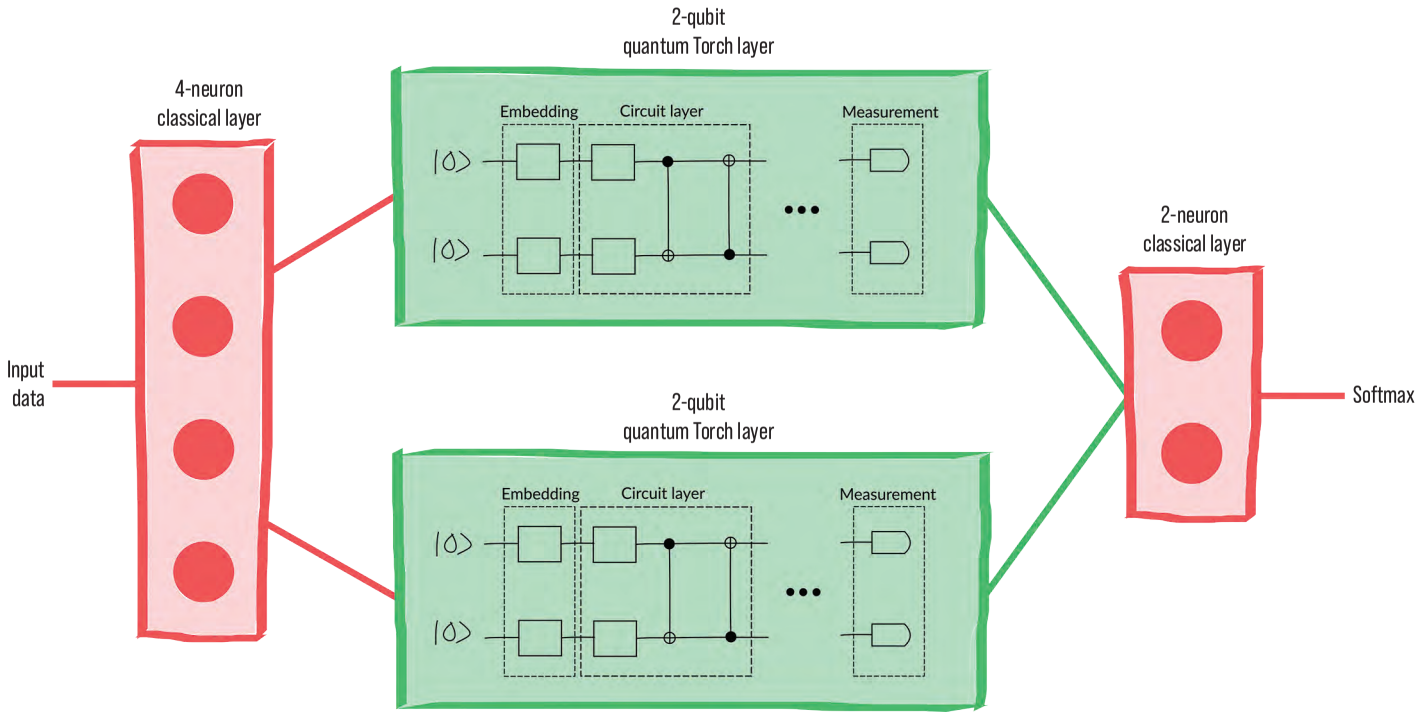
Źródło: Bromley [2024].

Rysunek 10.5. Schemat kwantowej warstwy sieci neuronowej z obwodem kodowania danych i obwodem wariacyjnym



Źródło: opracowanie własne.

Rysunek 10.6. **Schemat hybrydowej kwantowej sieci neuronowej z klasyczną warstwą wejściową, dwoma warstwami kwantowymi, zakończoną warstwą wyjściową z funkcją aktywacji softmax**



Źródło: Bromley [2024].

Podsumowanie

W niniejszej pracy przedstawione zostały podstawy matematyczne modeli kwantowego uczenia maszynowego oraz ich zastosowanie w problemie klasyfikacji binarnej. Przeanalizowano hybrydowe podejście bazujące na obwodach kwantowych i klasycznych optymalizatorach, realizujących warstwy ukryte klasycznej sieci neuronowej. Metody wariacyjnych algorytmów kwantowych mogą być wykorzystywane w problemach predykcyjnych oraz dla generowania nowych danych, a także w problemach optymalizacyjnych. Mimo że wciąż nie ma niezawodnego komputera kwantowego (duża liczba kubitów z długim czasem kontroli bramek oraz pełną korekcją błędów), pozwalającego w pełni stosować algorytmy kwantowe, metody hybrydowe mogą realizować w uproszczony (liniowy) sposób zadanie skomplikowanych sieci neuronowych.

Bibliografia

Benedetti, M., Lloyd, E., Sack, S., Fiorentini, M. (2019). Parameterized Quantum Circuits as Machine Learning Models, *Quantum Science and Technology*, 4(4), 043001.

Bromley, T. (2024). *Turning Quantum Nodes Into Torch Layers*, https://pennylane.ai/qml/demos/tutorial_qnn_module_torch (dostęp: 7.10.2024).

Cerezo, M., Sone, A., Volkoff, T. Cincio, L., Coles, P.J. (2021). Cost Function Dependent Barren Plateaus in Shallow Parametrized Quantum Circuits, *Nature Communications*, 12, 1791.

Cybulski, J.L., Zając, S. (2024). Design Considerations for Denoising Quantum Time Series Autoencoder. W: *Computational Science – ICCS 2024. ICCS 2024. Lecture Notes in Computer Science* (vol. 14837), L. Franco, C. de Mulatier, M. Paszyński, V.V. Krzhizhanovskaya, J.J. Dongarra, P.M. A Sloot (Eds.). Cham: Springer.

Deutsch, D., Jozsa, R. (1992). Rapid Solution of Problems by Quantum Computation, *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, 439(1907), s. 553–558.

Farhi, E., Goldstone, J., Gutmann, S. (2014.) *A Quantum Approximate Optimization Algorithm*. DOI: 10.48550/arXiv.1411.4028.

Feynman, R.P. (2022). *Lectures on Computation.* Boston: Addison-Wesley.

Gershenfeld, N.A., Chuang, I.L. (1996). Bulk Spin-Resonance Quantum Computation, *Science*, 275(5298), s. 350–356.

Grover, L.K. (1996). A Fast Quantum Mechanical Algorithm for Database Search. W: *STOC '96: Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing* (s. 212–219). New York: Association for Computing Machinery.

Harrow, A.W., Montanaro, A. (2017). Quantum Computational Supremacy, *Nature*, 549, s. 203–209.

Das, A., Chakrabarti, B. (2005). *Quantum Annealing and Related Optimization Methods*. Berlin: Springer-Verlag. DOI: 10.1007/11526216.

Havlíček, V., Mohseni, M., Lloyd, S. (2019). Supervised Learning with Quantum-Enhanced Feature Spaces, *Nature*, 567(7747), s. 209–212.

Hsu, C.-W., Ladd, T.D. (2020). Quantum Support Vector Classification, *Quantum*, 4, s. 328.

LaRose, R., Coyle, B. (2020). Robust Data Encodings for Quantum Classifiers, *Physical Review A*, 102(3), 032420.

Lund A.P., Bremner M.J., Ralph, T.C. (2017). Quantum Sampling Problems, Boson Sampling and Quantum Supremacy, *npj Quantum Information*, 3, 15.

Nayak, Ch., Simon, S.H., Stern, A., Freedman, M., Das Sarma, S. (2008). Non-Abelian Anyouns and Topological Quantum Computation, *Reviews of Modern Physics*, 80(3), s. 1083–1159.

Neumann von, J. (1945). *First Draft of a Report on the EDVAC*. Philadelphia: Moore School of Electrical Engineering, University of Pennsylvania.

Nielsen, M.A., Chuang, I. (2010). *Quantum Computation and Quantum Information*. Cambridge: Cambridge University Press.

Peruzzo, A., McClean, J., Shadbolt, P., Yung, M., Zhou, X., Love, P.J., Aspuru-Guzik, A., O'Brien, J.L. (2014). A Variational Eigenvalue Solver on a Photonic Quantum Processor, *Nature Communications*, 5(1), s. 4213.

Preskill, J. (2018). Quantum Computing in the NISQ Era and Beyond, *Quantum*, 2(79).

Przanowski, K, Zając, S. (2020). *Modelowanie dla biznesu. Metody machine learning, modele portfela consumer finance, modele rekurencyjne, analiza przeżycia, modele scoringowe*. Warszawa: Oficyna Wydawnicza SGH.

Schuld, M., Bocharov, A., Svore, K.M., Wiebe, N. (2020). Circuit-Centric Quantum Classifiers, *Physical Review A*, 101(3), 032308.

Shor, P.W. (1997). Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer, *SIAM Journal on Computing*, *26*(5), s. 1484–1509.

Sim, S., Johnson, P.D., Aspuru-Guzik, A. (2019). Expressibility and Entangling Capability of Parameterized Quantum Circuits for Hybrid Quantum-Classical Algorithms, *Advanced Quantum Technologies*, 2(12).

Turing, A. M. (1950). Computing Machinery and Intelligence, *Mind*, 59(236), s. 433–460.

Wong, T. G. (2022). *The Structure of Computer Systems*. Cambridge: Cambridge University Press.

Zając, S. (2022). *Modelowanie dla biznesu. Analityka w czasie rzeczywistym. Narzędzia informatyczne i biznesowe*. Warszawa: Oficyna Wydawnicza SGH.

III

REGULACJE I SPOŁECZNE KONSEKWENCJE WDRAŻANIA SZTUCZNEJ INTELIGENCJI

11.0

PODEJŚCIE DO REGULACJI STOSOWANIA SYSTEMÓW SZTUCZNEJ INTELIGENCJI – ANALIZA PORÓWNAWCZA WYBRANYCH KRAJOWYCH SYSTEMÓW PRAWNYCH

Aleksandra Szulc
Uniwersytet Ekonomiczny w Krakowie

Streszczenie

Tematyką rozdziału jest projektowanie regulacji dla sztucznej inteligencji (AI). Za cel postawiono przedstawienie wniosków z analizy porównawczej ram prawnych dla AI, w tym dyferencji między reżimami prawnymi, w ramach różnych podejść, przyjętych przez wybrane państwa, oraz udzielenie odpowiedzi na pytanie, czy możliwe jest wskazanie optymalnego rozwiązania prawnego. Badanie obejmuje aktywność legislacyjną przy zastosowaniu różnych narzędzi regulacyjnych w wybranych państwach, których selekcja została podporządkowana ukazaniu różnych perspektyw w regulowaniu AI. W opracowaniu została przedstawiona perspektywa nauk prawnych przy zastosowaniu studiów teoretycznych, metod dogmatyczno-prawnej i prawno-porównawczej. W dyskursie zostały wykorzystane pozycje literatury naukowej, dokumenty strategiczne, uchwalone akty normatywne, zbiory rekomendacji i wytycznych, a także raporty i opracowania przedstawicieli biznesu.

Słowa kluczowe: sztuczna inteligencja, otoczenie prawno-regulacyjne, analiza porównawcza, godna zaufania AI, soft law, efekt brukselski

Wprowadzenie

Wzrost zainteresowania podmiotów publicznych innowacjami na rynku i popularności rozwiązań technologicznych, wprowadzanych w sferze prywatnej, wzmocnił pilną potrzebę wprowadzenia przez regulatora zmian ich otoczenia prawnego. Nadążanie za transformacją cyfrową i ciągłym postępem technologicznym to niezbędny element nowoczesnego procesu stanowienia prawa. Pojawia się potrzeba ponownego spojrzenia na rolę państwa jako regulatora we współczesnym świecie oraz rewizja tradycyjnych paradygmatów i podejść w tworzeniu ram prawnych. Regulacje *ex post* w dobie wzmożonego rozwoju nowych technologii szybciej ulegają przedawnieniu, pozostawiając często martwe przepisy nieznajdujące zastosowania, dodatkowo bezrefleksyjne namnażanie aktów prawnych stanowi wyłącznie barierę dla dynamicznego rynku innowacji. Interwencjonizm legislatora powinien zostać ograniczony do sytuacji, w których obowiązujące prawo i samoregulacja rzeczonego obszaru okazują się niewystarczające dla zapewnienia prawidłowego funkcjonowania rynku [Flisak, 2019, s. 197]. Zmienność i wielowymiarowość nowych technologii, a przy tym częstokroć brak wiedzy specjalistycznej regulatorów sprawiają, że praca kreowania przepisów prawnych jest procesem ciągłym, wymagającym dostosowania narzędzi regulacyjnych, które skuteczniej i na dłuższy czas zagwarantują pewność prawa.

Tworząc regulacje, legislator zasadniczo dąży do przyjęcia rozwiązań wykorzystujących potencjał nowych technologii przy zagwarantowaniu ochrony nadrzędnych interesów, fundamentalnych praw, wartości i zasad demokratycznych. Inną barierę stanowią przenikanie się i interferencja dziedzin w ramach obszaru nowych technologii, co częstokroć rodzi potrzebę odmiennego uwarunkowania ich otoczenia prawnego [Chałubińska-Jentkiewicz, 2019, s. 54]. Ta interdyscyplinarność wiąże się z szerokim spektrum przedmiotu tej względnie nowej gałęzi prawa, w skład którego wchodzą ochrona danych osobowych, prawo własności intelektualnej, ochrona konkurencji i konsumentów oraz inne obszary zależne od platformy zastosowania innowacji.

11.1. Tworzenie otoczenia prawnego dla AI

Sztuczną inteligencję, zgodnie z definicją OECD, należy rozumieć jako „system oparty na maszynie, który może, dla danego zestawu celów zdefiniowanych przez człowieka, tworzyć prognozy, zalecenia lub podejmować decyzje wpływające na środowiska rzeczywiste lub wirtualne. Systemy AI są zaprojektowane do działania na różnych poziomach autonomii" [OECD, 2022]. Trudności z kreowaniem ram prawnych dla AI

wiążą się z ustalaniem zakresu regulacji, techniką legislacyjną, narzędziem regulacyjnym, a także podmiotem regulującym. Istotną przeszkodą jest częstokroć niska elastyczność prawa, objawiająca się szczególnie jako brak odporności na ich „zestarzenie się" w obliczu przyszłych zmian w innowacjach i szybki postęp technologiczny [Ard, Crootof, 2024, s. 34]. Narzędzia z obszaru twardego prawa mogą nie wytrzymać próby czasu, podczas gdy miękkie kodeksy zasadniczo automatycznie dostosują się do rynku. Coraz większą popularnością cieszą się piaskownice regulacyjne, w ramach eksperymentalnego podejścia do prawa, przeznaczone do testowania nowych rozwiązań technologicznych w bezpiecznych warunkach i przy derogacji części przepisów prawa. Wielowymiarowość, zmienność, ciągły rozwój i samouczenie się maszyny wymagają od prawodawcy elastycznego podejścia przy jednoczesnym ustanowieniu chociażby minimalnych fundamentów prawno-regulacyjnych [Dziedzic, 2022, s. 350]. Pojawienie się tej przełomowej innowacji i jej etapowa absorpcja przez rynek narzuciła na ustawodawcę podjęcie działań w celu zagwarantowania ochrony praw jego uczestników i zniwelowania ryzyka materializacji negatywnych skutków. Trwałość ustanowionego porządku prawnego ugruntowanego na wartościach powinna zostać zachowana, a wszelkie konflikty związane z wyzwaniami technologicznymi zażegnane przez państwo [Weber, 2021, s. 55] Tym samym prawodawca jako gwarant stabilności powinien podjąć adekwatne działania na podłożu prawno-regulacyjnym.

W ostatnich latach krajowi i ponadnarodowi prawodawcy uznali, że rynek sztucznej inteligencji osiągnął dojrzałość na tyle, że zamknięcie obszarów wykorzystania AI w ramy prawne stało się kwestią wymagającą aktywności regulacyjnej. Zasadniczo regulator dokonuje wyboru pomiędzy trzema różnymi ścieżkami. Pierwsze podejście polega na wykorzystaniu rozumowania *per analogiam* i wykładni rozszerzającej do objęcia tych nowości obowiązującymi, neutralnymi pod względem technologicznym, ramami prawnymi [Ard, Crootof, 2024, s. 29]. Kolejne rozwiązanie koncentruje się wokół tworzenia nowego reżimu, często o charakterze *ex ante* dla innowacji, których jeszcze nie zaaplikowano na rynku. Ostatnia ścieżka jest domeną nadzoru. Instytucje ograniczają swoje działania w zakresie otoczenia prawno-regulacyjnego do kwestii udzielania zezwoleń, licencji i rejestracji, wprowadzonych do obrotu nowoczesnych rozwiązań, i ich dostawców. Przyjmując inną perspektywę, część systemów prawnych podążyło drogą twardego prawa, projektując i wdrażając akt normatywny o randze ustawowej, czy jako akt wykonawczy (administracyjny) regulujący narzędzia wykorzystujące AI. Odmienne podejście polega na zastosowaniu mechanizmów *soft law*, takich jak kodeksy postępowania, wytyczne i rekomendacje, często przy wsparciu działań rządowych w postaci polityk, strategii i odgórnych ram [Chan, Papyshev, Yarime, 2022, s. 2].

11.2. Sztuczna inteligencja w obszarze zainteresowania prawodawców

Wiele systemów prawnych podjęło działania, które można określić jako „wczesny etap" tworzenia ram prawnych dla AI. Zasadniczo rozpoczyna się od poświęcenia przez prawodawcę szczególnej uwagi rozwiązaniom prawnym w obszarze przetwarzania danych, których jakość i bezpieczeństwo stanowią fundament dla realizacji godnej zaufania AI [Harasimiuk, Braun, 2021, s. 74]. Dane są najistotniejszym elementem AI i na nich opiera się cały proces algorytmicznego przetwarzania, niezależnie od stopnia zaawansowania systemu [Nowakowski, 2023, s. 46]. Nie zaskakuje więc fakt, iż wpierw musiały zostać postawione solidne fundamenty pod *data governance*. Uchwalone w UE rozporządzenie o ochronie danych osobowych (RODO) stało się inspiracją dla aktualizacji i unowocześnienia tych ram w różnych krajowych systemach prawnych. Wpływ wskazanej regulacji stanowi przykład występowania tzw. efektu brukselskiego (*Brussels effect*). Zjawisko to polega na transponowaniu reżimów prawnych UE do pozaunijnych systemów prawnych poprzez mechanizm rynkowy i zaadaptowanie ich jako standardów [Li i in., 2023, s. 23]. W kontekście przedmiotowej problematyki szczególną uwagę poświęcono kwestii oparcia decyzji wyłącznie na zautomatyzowanym przetwarzaniu danych i profilowaniu, co zostało uregulowane w art. 22 RODO.

Impulsem dla intensyfikacji działań w ramach tworzenia otoczenia prawno-regulacyjnego dla AI na skalę międzynarodową było opracowanie przez grupę ekspertów, a następnie przyjęcie 22 maja 2019 r. przez Radę OECD zbioru rekomendacji w zakresie etycznej i godnej zaufania AI. Mimo niewiążącego charakteru dokument przez kolejne lata pełnił i pełni nadal rolę powszechnie akceptowalnego wzorca dla standardów rozwoju sztucznej inteligencji [Lim, 2022, s. 93]. Wytyczne zawierają rekomendacje dotyczące opracowywania krajowej polityki i opierają się na zasadach: zrównoważonego rozwoju, uczciwości, skoncentrowania na człowieku, przejrzystości, wyjaśnialności, solidności, ochrony i bezpieczeństwa czy rozliczalności [OECD, 2022]. Ten przełomowy dokument nieznacznie wyprzedziła grupa ekspertów powołana przez Komisję Europejską w czerwcu 2018 r. Zespół opracował i przedstawił na szczeblu UE wytyczne w zakresie etyki, dotyczące godnej zaufania sztucznej inteligencji, czyli zgodnej z prawem, etycznej i solidnej. Na pierwszy plan wysunięto zasady poszanowania autonomii człowieka, prewencji negatywnych skutków, sprawiedliwości i wyjaśnialności [Grupa Ekspertów Wysokiego Szczebla do spraw Sztucznej Inteligencji, 2019].

Rozbudowane zalecenia dla swoich członków, dotyczące etycznego obszaru stosowania narzędzi AI, uchwaliło UNESCO podczas 41. sesji Konferencji Generalnej w listopadzie 2021 r. Poza wskazaniem wartości i zasad wyznaczających ramy prawne dla AI w rekomendacjach można odnaleźć odwołania do obszarów, takich jak edukacja,

nauka, kultura, komunikacja oraz do problematyki równości płci, ochrony środowiska i ekosystemów [UNESCO, 2021].

Omawiane dokumenty charakteryzują się dość szczątkową dyskusją na temat przyszłych ram prawnych, skupiając się w swoim przekazie na podkreśleniu wagi aspektu etycznego [Mantelero, 2022, s. 47]. Etyczność, przejrzystość oraz zgodność z prawem i wartościami powinny być, zgodnie z tymi wytycznymi, wpisane w każdy etap cyklu życia systemów sztucznej inteligencji. Zbiory wytycznych skoncentrowane na człowieku, godnej zaufania i odpowiedzialnej AI niejednokrotnie inkorporowano do krajowych porządków prawnych. Wielokrotnie odwołują się do nich rządowe strategie wyznaczające podejścia i kierunki w zakresie budowania otoczenia prawno-regulacyjnego dla AI, w tym przygotowania podatnego i bezpiecznego gruntu dla zastosowań tej technologii na rynku i w różnych obszarach aktywności państwa.

11.3. Podejście holistyczne oparte na analizie ryzyka

W podejściu holistycznym wysiłki regulacyjne są skoncentrowane na opracowaniu ram prawnych, stanowiących kompleksowe ujęcie problemu miejsca sztucznej inteligencji w systemie prawa. Niejednokrotnie aktywność ustawodawcy obejmuje nowelizację obowiązujących regulacji, szczególnie w zakresie cyfryzacji, cyberbezpieczeństwa i ochrony danych, tym samym rozpraszając przepisy dotyczące AI w różnych aktach normatywnych. Często legislator wybiera pewien obszar, sektor, kluczowy obszar bądź narzędzie AI wymagające jego interwencji. Do względnej rzadkości należy całościowe ujęcie zastosowania AI w ramach jednego międzysektorowego aktu prawnego, obejmującego wszystkie możliwe obszary zastosowania tych systemów.

Zasadniczo wykształciły się dwa odrębne podejścia w opracowaniu regulacji sztucznej inteligencji, przy czym można dodatkowo wyróżnić stanowisko mieszane. Pierwsze sprowadza się do ujęcia przepisów prawnych dotyczących AI w oparciu na analizie ryzyka. Punktem centralnym regulacji jest człowiek i wpływ, jaki sztuczna inteligencja może na niego wywierać. Najważniejszymi reprezentantami tego nurtu jest Unia Europejska i Kanada, w których wysiłki legislacyjne skupiono na opracowaniu międzysektorowego i kompletnego aktu prawnego o randze ustawowej. Drugie podejście ma na celu położenie fundamentów prawnych sprzyjających innowacjom i rozwojowi rynku. Przykładami systemów prawnych najbardziej odpowiadających tym założeniom dysponują Stany Zjednoczone i Chiny.

Unia Europejska, podejmując się wyzwania zaprojektowania regulacji harmonizującej zasady wykorzystywania systemów AI, postawiła sobie za cel realizację idei

godnej zaufania sztucznej inteligencji, zgodnie z którą na pierwszym miejscu stawia się człowieka i ochronę jego praw na czele z godnością ludzką (podejście antropocentryczne) [Flisak, 2024]. To główne założenie przyjęto jako bazę w unijnej strategii w zakresie sztucznej inteligencji, a następnie w komunikacie Komisji Europejskiej z 8 kwietnia 2019 r. i białej księdze „Europejskie podejście do doskonałości i zaufania". W unijnym podejściu posłużono się koncepcją *ethics by design*, zgodnie z którą aspekt etyczny uwzględniany jest już na etapie projektowania reżimu prawnego [Harasimiuk, Braun, 2021, s. 54]. Rezultatem kilkuletnich prac był projekt rozporządzenia, który miał na celu pogodzić dążenia do budowania konkurencyjnego jednolitego rynku i realizacji koncepcji godnej zaufania AI przy jednoczesnym osiągnięciu równowagi pomiędzy innowacyjnością a ochroną uczestników obrotu gospodarczego [Keller, Pereira, Pires, 2024, s. 428].

Akt w sprawie sztucznej inteligencji stanowi zbiór międzysektorowych przepisów, harmonizujących zasady wykorzystywania systemów AI, zgodnie z podejściem opartym na analizie ryzyka. Systemy o niedozwolonym poziomie ryzyka, np. służące scoringu społecznemu czy manipulacji podprogowych (Motyw 29 i 31), zebrano w katalog zakazanych praktyk i zastosowań, zaś w przypadku tych o niskim stopniu ryzyka za wystarczające uznano spełnienie wymogu przejrzystości (Motyw 26). Systemy AI wysokiego ryzyka, klasyfikowane na podstawie ich potencjalnej szkodliwości dla praw podstawowych (Motyw 48), mogą zostać dopuszczone na rynek i oddane do użytku po spełnieniu przez dostawców szeregu obowiązków, od oceny *ex ante* zgodności z przepisami unijnymi, po zarządzanie ryzykiem w całym cyklu jego życia.

Organy unijne w tworzeniu reżimu prawnego AI posłużyły się rozporządzeniem, czyli aktem normatywnym, zakładającym bezpośrednie stosowanie w państwach członkowskich. Sama regulacja przewiduje bieżące monitorowanie i powierzenie krajowym organom nadzorczym sprawowania kontroli nad wdrażaniem zunifikowanych przepisów zgodnie z ustalonym harmonogramem, a także uznaniową implementację miękkich zasad, w postaci kodeksów postępowania czy samoregulacji rynkowych.

Analogiczne podejście oparte na analizie ryzyka odzwierciedla propozycja Artificial Intelligence and Data Act, stanowiąca trzecią część projektu C-27, procedowaną w kanadyjskim parlamencie. Celem regulacji jest harmonizacja wymogów dla projektowania, rozwoju i użytkowania systemów AI oraz wprowadzenie katalogu praktyk zakazanych ze względu na ich potencjalne negatywne i zarazem istotne skutki dla osób fizycznych i ich interesów [AIDA, 2023, art. 4]. W przeciwieństwie do wcześniejszych inicjatyw na poziomie federalnym, skupionych wokół rozwoju badań, innowacji i samego rynku AI, punkt koncentracji projektu regulacji został położony

na mitygacji ryzyka i minimalizacji potencjalnych szkód, związanych z zastosowaniem tych systemów [CLC, 2023, s. 2]. W AIDA odpowiednikiem systemu wysokiego ryzyka z unijnego rozporządzenia jest „hight-impact AI system", czyli system o znaczącym wpływie. Samej definicji tych systemów nie odnajdziemy w projekcie, a jedynie odwołanie do przyszłych rozporządzeń, których uchwalenie powierzono gubernatorowi w radzie (Governor in Council, GIC) [AIDA, 2023, art. 36]. Zasadniczo delegowanie zadań w zakresie uregulowania czy doprecyzowania dotyczy dość szerokiej grupy terminów i obowiązków wskazanych w AIDA. Sam zakres obowiązywania zasad wykorzystywania narzędzi AI obejmuje wyłącznie sektor prywatny. Na poziomie federalnym zastosowaniu podlega dyrektywa w sprawie zautomatyzowanego podejmowania decyzji (*Directive on Automated Decision-Making*) z 2019 r. oraz szereg narzędzi i instrumentów opublikowanych przez instytucje federalne [Attard-Frost i in., 2024, s. 7]. Opracowanie kompleksowej i międzysektorowej regulacji dla sztucznej inteligencji wiąże się z szeregiem wyzwań na poziomie prawnym. Jako pierwszą barierę należy wymienić podział kompetencji. Prawo w obszarze konkurencji powierzono organom federalnym, ochrony konsumentów – instytucjom prowincjonalnym, zaś dziedzina ochrony danych, prywatności i praw człowieka wymaga współpracy obu sektorów administracji publicznej [Martin-Bariteau, Scassa, 2021, s. 9]. Wypracowanie kompromisu utrudnia także konieczność ingerencji w domenę dotychczas przynależną do różnych agencji i departamentów wykorzystujących narzędzia typu *soft law* [Martin-Bariteau, Scassa, 2021, s. 9].

Analizując historię aktywności legislacyjnej w obszarze AI w Brazylii, nie w sposób nie zauważyć licznych projektów aktów normatywnych, ich modyfikacji i konsolidacji w kolejne propozycje ram prawnych. Pierwszym projektem, który spotkał się z największym zainteresowaniem interesariuszy i jednocześnie krytyką, był projekt 21/2020, oparty na zasadach godnej zaufania AI i zakładający wymóg przejrzystości w stosowaniu i projektowaniu rozwiązań AI [Belli, Curzi, Gaspar, 2023, s. 9]. Minimalistyczna i dość ogólnikowa propozycja odzwierciedlała interesy korporacji i miała stanowić podatny grunt dla rozwoju zdigitalizowanego rynku przy znikomym odniesieniu się do ochrony praw podstawowych [Keller, Magalhães, 2023, s. 189–190]. Przedmiotem aktualnych dyskusji w brazylijskim parlamencie jest projekt 2338/2023, który czerpie garściami z unijnego wzorca. W projekcie przyjęto podejście oparte na analizie ryzyka, uzależniając wymagane standardy i zakres obowiązków od stopnia ryzyka związanego z danym systemem AI oraz koncentrując się na ochronie praw i interesów osób dotkniętych efektami działania sztucznej inteligencji [Frazão, 2024, s. 56].

11.4. Podejście rynkowe w ustawodawstwie w obszarze AI

Reprezentantem drugiego podejścia w obszarze twardego prawa, ukierunkowanego na budowanie konkurencyjnego rynku AI, są Chiny. Liczący się gracze na chińskim rynku technologii to duże przedsiębiorstwa, których zadaniem jest rozwój poszczególnych obszarów AI przy wsparciu rządowym [Roberts, Cowls, Morley, Taddeo, Wang, Floridi, 2021, s. 61]. Państwo przy użyciu narzędzi regulacyjnych i administracyjnych przygotowuje podatny grunt pod ekspansję tych firm przy jednoczesnym kontrolowaniu działań monopolistycznych i utrzymywaniu porządku społecznego [Filipova, 2024, s. 55]. Nadrzędnym celem rządu jest objęcie pozycji lidera w budowaniu innowacyjnego rynku dla AI. Przedmiot zainteresowania Chin stanowi problematyka wpływu nowych technologii na społeczeństwo jako całość, zaś troska o indywidualne interesy i ochronę jednostek, która jest silnie widoczna w unijnym modelu, ma tu niewielkie znaczenie [Rolf, 2023, s. 5]. Chiny jako pierwsze wdrożyły ramy prawne dla wybranych obszarów zastosowania systemów AI. W marcu 2022 r. weszły w życie przepisy administracyjne, wprowadzające obowiązki dla platform internetowych świadczących usługi w zakresie generowania rekomendacji w oparciu na algorytmach [China Law Translate, 2022b]. W tym samym roku opublikowano akt normatywny dotyczący przedmiotu głębokiej syntezy, w którym uregulowano zakres odpowiedzialności dostawców i obsługi technicznej za bezpieczeństwo danych, ochronę danych osobowych, przejrzystość i zarządzanie systemami AI wykorzystywanymi w świadczeniu usług oraz mechanizmy zwalczania fałszywych treści (tzw. *deep fake*) [China Law Translate, 2022a].

Popularna koncepcja godnej zaufania i odpowiedzialnej AI znalazła swoich zwolenników także w Chinach, na co wskazuje wyrażenie przez Ministerstwo Spraw Zagranicznych Chin potrzeby priorytetyzacji etycznego obszaru sztucznej inteligencji w dokumencie „Position Paper of the People's Republic of China on Strengthening Ethical Governance of Artificial Intelligence (AI)", opublikowanym 17 listopada 2022 r. Zasady etycznej AI zostały szeroko rozpowszechnione w ramach samoregulacji w różnych sektorach, czego przykładem jest zbiór wytycznych dla obszaru badań i rozwoju instytutu badawczego „Beijing AI Principles".

Koncentrację na rozwoju rynku AI i budowaniu konkurencyjnej gospodarki można dostrzec w projektach legislacyjnych w Tajwanie i Korei Południowej. W trzech propozycjach Basic Law on Artificial Intelligence, datowanych na lata 2019–2022, wykorzystano wzorce amerykańskie, japońskie i UE w próbach uregulowania aspektów projektowania, wdrażania i stosowania systemów AI w Tajwanie [Hsu, 2024, s. 151–152]. Ostatni projekt regulacji został oparty na założeniu partnerstwa publiczno-prywatne-

go, zgodnie z którym podmioty gospodarcze wdrożą normatywne zasady rozwoju AI, a zadania rządowe zostaną skoncentrowane na promocji dalszych postępów w obszarze tej technologii [Hsiung, Sung, 2024]. Podobne stanowisko zajął rząd Korei Południowej, który jeszcze niedawno wiernie podążał ścieżką obraną przez państwa Azji Wschodniej, opartą na miękkich regulacjach w ramach etycznej i godnej zaufania AI [Singh, 2023, s. 119]. Z propozycji ram prawnych dla AI największy rozgłos zyskał projekt „Act on Promotion of AI Industry and Framework for Establishing Trustworthy AI" [Ko, 2024]. Pomimo podobieństw do unijnego rozporządzenia w obszarze różnicowania obowiązków dostawców w zależności od stopnia ryzyka czy wyróżnienia systemów AI wysokiego ryzyka, koreańska propozycja ma na celu przede wszystkim wzmocnienie konkurencyjności rynku i przygotowanie chłonnego gruntu pod innowacje [Ko, 2024]. W polityce regulacyjnej tych państw widoczne są zmagania pomiędzy europejskim a amerykańskim wpływem, co wiąże się z zainteresowaniem azjatyckich gospodarek udziałem w obu tych rynkach [An, 2024, s. 1].

11.5. Podejście do regulowania AI oparte na zasadach i narzędziach regulacyjnych typu *soft law*

Państwa, które podążyły rzeczoną ścieżką, mają w świadomości kompleksowość i zależność wpływu i potencjalnych skutków zastosowania narzędzi AI od danego sektora, którego specyfika i unikatowe potrzeby wymagają dostosowanego podejścia regulatora [Walter, 2024, s. 2]. Prawodawcy, nadzorcy i administracja rządowa koncentrują swoje działania na miękkim zarządzaniu i kontroli podmiotów gospodarczych, które wykorzystują lub planują wdrożyć systemy AI. Elementem charakteryzującym to podejście jest opracowanie zbioru wytycznych i zasad, wzorowane na wspomnianych wcześniej ponadnarodowych dokumentach w obszarze etycznej AI.

Rozważania warto rozpocząć od analizy brytyjskiego podejścia do regulowania sztucznej inteligencji. Zasadniczo jego celem jest tworzenie reżimu prawnego sprzyjającego innowacjom, spełniającego warunek proporcjonalności, wykazującego elastyczność w ciągłym dostosowywaniu się do zmian technologicznych i opartego na zasadach odpowiedzialnej AI [Department for Science, Innovation and Technology, 2023, s. 6]. Organy regulacyjne w zakresie swoich kompetencji konstruują ramy prawne, wykorzystując głównie narzędzia regulacyjne typu *soft law* dla rozwiązań AI, znajdujących zastosowanie w regulowanych domenach [Roberts, Babuta, Morley, Thomas, Taddeo, Floridi, 2023, s. 3]. Przykładem są wytyczne Biura Komisarza ds. Informacji [ICO, 2024] w obszarze wdrażania sztucznej inteligencji przy zachowaniu zgodności z systemem

ochrony danych osobowych („Guidance on AI and Data Protection") i w przedmiocie wyjaśniania decyzji, wydanych przy użyciu systemów AI („Explaining Decisions Made with AI") [ICO, 2024]. Zgodnie z omawianym sektorowym podejściem rząd brytyjski może w ramach swoich uprawnień wydawać niewiążące zalecenia skierowane do regulatorów, zawierające pożądane kierunki rozwoju otoczenia prawnego dla nowych technologii. Przykładem są wstępne wytyczne Department for Science, Innovation and Technology, opublikowane w lutym 2024 r.

Szereg podobieństw do brytyjskiego podejścia można odnaleźć w japońskim stanowisku wobec tworzenia ram prawnych dla systemów AI. Stanowisko Kraju Kwitnącej Wiśni sprowadza się do publikacji niewiążących wytycznych dopasowanych do specyfiki danego sektora, dodatkowo pozostawiając szerokie pole dla samoregulacji [Walter, 2024, s. 9]. Pierwszym kluczowym dokumentem jest „Social Principles of Human-Centric AI", zawierający zbiór zasad: skoncentrowania na człowieku, edukacji, ochrony prywatności, zapewnienia bezpieczeństwa, uczciwej konkurencji, sprawiedliwości, rozliczalności i przejrzystości oraz innowacji, które mają wytyczać rozwój AI. Odbiorcą wytycznych jest całe społeczeństwo, w tym deweloperzy i operatorzy na rynku AI [Expert Group on How AI Principles Should be Implemented, 2022, s. 3]. Drugi zbiór, „AI Governance Guidelines", zawiera cele działań wspierające prawidłowe wdrożenie zasad związanych z systemami sztucznej inteligencji z elementami kazuistyki, ułatwiającymi zidentyfikowanie luk i podjęcie czynności dostosowawczych [Expert Group on How AI Principles Should be Implemented, 2022, s. 3]. Japońskie podejście opiera się na elastyczności i ciągłej aktualizacji otoczenia prawno-regulacyjnego. Publikowane wytyczne nie wprowadzają żadnych zakazów dotyczących praktyk czy rozwiązań technologicznych, a wdrożenie odpowiednich narzędzi mitygujących ryzyko (podejście oparte na analizie ryzyka) i określenie zakresu obowiązków pozostawia się w gestii podmiotów gospodarczych w ramach tzw. *agile governance* [Habuka, 2023, s. 3].

Innym przykładem państwa, które postawiło na miękkie, wspierające rynek i jego uczestników narzędzia regulacyjne, jest Singapur. Personal Data Protection Commission (PDPC) opublikowała w styczniu 2019 r. modelowe ramy („Model AI Governance Framework") stanowiące międzysektorowy zbiór zasad w zakresie projektowania, wdrażania i użytkowania ogólnie pojmowanej AI w obszarach zarządzania wewnętrznego, zarządzania operacyjnego, zarządzania relacjami z klientami i procesów decyzyjnych. Innym przykładem instrumentu *soft law* są wytyczne odnoszące się do przetwarzania danych osobowych przez systemy AI, generujące rekomendacje, predykcje i decyzje, które zostały opublikowane 1 marca 2024 r. przez PDPC. Za szczególnie wyróżniający się na tle innych zaleceń należy uznać zbiór zasad FEAT (Fairness, Ethics, Accountability and Transparency), czyli sprawiedliwości, etyki, rozliczalności i przejrzystości, wyda-

ny przez Monetary Authority of Singapore (MAS) dla narzędzi AI i analityki danych w sektorze finansowym. Wskazane rekomendacje służą praktycznemu zaaplikowaniu przez przedsiębiorstwa ram dla odpowiedzialnego wdrażania sztucznej inteligencji, realizacji zasady pewności prawa i zagwarantowaniu sprzyjającego gruntu pod innowacje [IMDA, PDPC, 2020, s. 4].

11.6. Podejście oparte na partnerstwie publiczno-prywatnym w Stanach Zjednoczonych

Po omówieniu w poprzednim podrozdziale wzoru europejskiego, chińskiego i modelu *soft law* dla kreowania ram prawnych, za przedmiot analizy przyjęto podejście, które łączy i czerpie z tych kierunków. Reprezentantem o dość rozbudowanej strukturze źródeł prawa i narzędzi regulacyjnych są Stany Zjednoczone. Amerykańskie otoczenie prawno-regulacyjne dla AI stanowi pluralistyczne i wielowymiarowe środowisko, złożone z ustawodawstwa rządu federalnego, prawa stanowego, dekretów prezydenckich, ogólnokrajowych wytycznych, zbiorów standardów, kodeksów postępowania i innych ram prawnych. Bazą podejścia Stanów Zjednoczonych do regulowania AI jest współregulacja publiczno-prywatna i zorientowanie na dany sektor. Odbiorcą aktów prawnych na poziomie federalnym są zasadniczo regulatorzy danych sektorów. Rolą tych regulacji jest wyznaczenie kierunków aktywności agencji i organów rządowych w obszarze rozwoju sztucznej inteligencji. Jako przykłady warto wskazać National Artificial Intelligence Initiative Act (NAIIA) of 2020 oraz The AI in Government Act of 2020. Później za punkt koncentracji przyjęto także same systemy AI i zasady ich wdrażania, czego dowodem jest Algorithmic Accountability Act of 2022, dotyczący wymogu przeprowadzenia analizy wpływu zautomatyzowanych procesów decyzyjnych na prawa i interesy konsumentów.

Wśród miękkich narzędzi regulacyjnych na wyróżnienie zasługują biała księga „Blueprint for an AI Bill of Rights" oraz *Artificial Intelligence Risk Management Framework*, wydany przez National Institute of Standards and Technology [Tabassi, 2023]. Pierwszy dokument stanowi zbiór niewiążących wytycznych, opartych na pięciu zasadach, których celem jest wsparcie podmiotów publicznych i prywatnych na etapie projektowania, użytkowania i wdrożenia systemów AI, zgodnie z ogólnie przyjętymi wartościami i prawami podstawowymi (AI Risk Management Framework). Tym, co wyróżnia te ramy prawne, jest skoncentrowanie na ochronie jednostek przed potencjalnymi negatywnymi skutkami, związanymi z zautomatyzowanymi procesami decyzyjnymi, oraz dostosowanie polityk i działań agencji rządowych do poszczególnych

sektorów [Engler, 2023]. Drugi dokument zawiera dobrowolne, międzysektorowe, elastyczne i niezależne od wielkości organizacji zasady w zakresie zarządzania ryzykiem związanym z systemami AI (AI Risk Management Framework). Obydwa zbiory wytycznych świadczą o zwrocie kierunku w projektowaniu regulacji w stronę etycznego wymiaru sztucznej inteligencji w ramach szeroko promowanej koncepcji godnej zaufania i odpowiedzialnej AI [Khan, Shoaib, Arledge, 2024, s. 5].

Podsumowanie

W procesie budowania otoczenia prawno-regulacyjnego AI niewątpliwie swoją obecność zaznaczyły zarówno Unia Europejska z podejściem opartym na wartościach wspólnotowych, harmonizacji prawa i analizie ryzyka, Stany Zjednoczone z sektorową koncentracją wielopoziomowych regulacji, Chiny z nastawieniem na dominację gospodarczą i szybki rozwój krajowego rynku, jak i Wielka Brytania z Japonią w roli partnerów dla przedsiębiorców. Te różnice wynikają z przyjętych celów polityki i realizujących je narzędzi regulacyjnych, a także z samych dyferencji systemowych w tworzeniu otoczenia prawnego. Stanowisko restrykcyjne i protekcjonistyczne wobec konsumentów bazuje często na twardym prawie, zaś w podejściu promującym innowacje zasadniczo korzysta się z rozwiązań *soft law* i dopuszcza się szeroką samoregulację. Każde z państw dąży do odegrania istotnej roli w tworzeniu nowoczesnego ekosystemu AI i zajęcia pozycji lidera na rynku tej technologii. Państwa będące świadkami tego wyścigu zasadniczo sięgają po wzorce promowane przez liderów, dobierając narzędzia regulacyjne i preferowane elementy różnych podejść do regulowania AI. Zwolennikami eksperymentalnego podejścia do prawa (piaskownice regulacyjne) są najczęściej reprezentanci podejścia opartego na partnerstwie prywatno-publicznym, w szczególności Wielka Brytania i Singapur.

Odpowiadając na postawione pytanie badawcze, nie można wskazać jednej optymalnej drogi dla uregulowania AI. Akt w sprawie sztucznej inteligencji został opracowany zgodnie z unijnymi wartościami i zasadami, które są względnie unikalne dla Wspólnoty Europejskiej, zaś chińskie i amerykańskie ramy prawne mają służyć rozwojowi ich gospodarek. Różnorodność kultur prawnych, metod tworzenia prawa, historii prawodawstwa, potrzeb rynku i jego uczestników uniemożliwia implementowanie z powodzeniem jednego wybranego wzorca. Dodatkowo specyfika każdego z obszarów aktywności państwa wymaga indywidualnego podejścia. Zbyt ogólnikowe ujęcie problematyki AI z oparciem wyłącznie na miękkich zasadach generuje ryzyko pozostawienia w porządku prawnym „martwych i pustych" przepisów, zaś sztywna

i restrykcyjna regulacja stanie się wyłącznie hamulcem dla rozwoju rynku i innowacji. Ponadto nadmierne zróżnicowanie ram prawnych przekłada się na trudności w obszarze *compliance,* wzmacnia zjawisko arbitralności regulacyjnej i tzw. *race to the bottom*. Wskazane zjawisko stanowi efekt wzmacniania konkurencyjności krajowego rynku celem przyciągnięcia inwestorów, szczególnie w ramach podejścia sektorowego do regulowania AI, kosztem obniżenia standardów i łagodzenia wymogów prawnych wobec systemów AI [Li, Schütte, Sankari, 2023, s. 22–23]. Rozwiązaniem może być wieloaspektowa współpraca transgraniczna, która ma szansę ostudzić gorączkę regulacyjną i wyznaczyć kierunek rozwoju rynku AI na najbliższe lata.

Bibliografia

An, J. (2024). The EU AI Act and Its Strategic Implications for South Korean AI Companies, *EAF Policy Debates*, 215, https://www.keaf.org/en/book/EAF_Policy_Debates/The_EU_AI_Act_and_Its_Strategic_Implications_for_South_Korean_AI_Companies?ckattempt=1 (dostęp: 31.07.2024).

Ard, B.J., Crootof, R. (2024). Legal Responses to Techlaw Uncertainties. W: *Research Handbook on Law and Technology* (s. 28–44), B. Brożek, O. Kanevskaia, P. Pałka (Eds.). Northampton: Edward Elgar Publishing.

Artificial Intelligence and Data Act of Bill C-27(2022). *An Act to Enact the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial Intelligence and Data Act and to Make Consequential and Related Amendments to Other Acts*, https://www.parl.ca/documentviewer/en/44-1/bill/C-27/first-reading (dostęp: 2.08.2024).

Attard-Frost, B., Brandusescu, A., Lyons, K. (2024). The Governance of Artificial Intelligence in Canada: Findings and Opportunities from a Review of 84 AI Governance Initiatives, *Government Information Quarterly*, 41(2), 101929.

Beijing Academy of Artificial Intelligence (2019). Beijing AI Principles, *Datenschutz und Datensich*, 43, s. 656. DOI: 10.1007/s11623-019-1183-6.

Belli, L., Curzi, Y., Gaspar, W.B. (2023). AI Regulation in Brazil: Advancements, Flows, and Need to Learn from the Data Protection Experience, *Computer Law and Security Review*, 48, 105767.

Chałubińska-Jentkiewicz, K. (2019). Rozwój nowoczesnych technologii w kontekście procesu stanowienia prawa na przykładzie strategii AI, *Teka Komisji Prawniczej PAN Oddział w Lublinie*, 12(2), s. 53–71.

Chan, K.J.D., Papyshev, G., Yarime, M. (2022). *Balancing the Tradeoff Between Regulation and Innovation for Artificial Intelligence: An Analysis of Top-down Command and Control and Bottom-Up Self-Regulatory Approaches*, https://ssrn.com/abstract=4223016 (dostęp: 3.08.2024).

Chik, W. (2021). The Future of Personal Data Protection Law in Singapore: A Role for the Use of AI and the Propertisation of Personal Data. W: *AI, Data and Private Law Translating Theory into Practice AI*, G.C.K. Yew, M. Yip (Eds.). Oxford: Hart Publishing.

China Law Translate (2022a). *Provisions on the Administration of Deep Synthesis Internet Information Services*, https://www.chinalawtranslate.com/en/deep-synthesis/ (dostęp: 2.08.2024).

China Law Translate (2022b). *Provisions on the Management of Algorithmic Recommendations in Internet Information Services*, https://www.chinalawtranslate.com/en/algorithms/ (dostęp: 2.08.2024).

CONGRESS.GOV (2020a). *H.R.2575 – AI in Government Act of 2020*, https://www.congress.gov/bill/116th-congress/house-bill/2575 (dostęp: 5.08.2024).

CONGRESS.GOV (2020b). *H.R.6216 – National Artificial Intelligence Initiative Act of 2020*, *116th Congress (2019–2020)*, https://www.congress.gov/bill/116th-congress/house-bill/2575 (dostęp: 5.08.2024).

CONGRESS.GOV (2022). *H.R. 6850 and S. 3572 – Algorithmic Accountability Act of 2022*, *117th Congress (2021–2022)*, https://www.congress.gov/bill/117th-congress/senate-bill/3572/ and https://www.congress.gov/bill/117th-congress/house-bill/6580/ (dostęp: 5.08.2024).

Department for Science, Innovation and Technology (2023). *A Pro-Innovation Approach to AI Regulation*, https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper (dostęp: 2.08.2024).

Department for Science, Innovation and Technology (2024). *Implementing the UK's AI Regulatory Principles: Initial Guidance for Regulators*, https://www.gov.uk/government/publications/implementing-the-uks-ai-regulatory-principles-initial-guidance-for-regulators (dostęp: 3.08.2024).

Dziedzic, M. (2022). Zastosowanie sztucznej inteligencji na rynkach finansowych w kontekście proponowanych zmian regulacyjno-prawnych. W: *Prawo sztucznej inteligencji i nowych technologii 2* (s. 347–366), B. Fischer, A. Pązik, M. Świerczyński (red.). Warszawa: Wolters Kluwer.

Engler, A. (2023). *The EU and U.S. Diverge on AI Regulation: A Transatlantic Comparison and Steps to Alignment*, https://www.brookings.edu/articles/the-eu-and-us-diverge-on-ai-regulation-a-transatlantic-comparison-and-steps-to-alignment/ (dostęp: 5.08.2024).

Expert Group on How AI Principles Should be Implemented (2022). *Governance Guidelines for Implementation of AI Principles, Ver. 1.1, AI Governance Guidelines WG*, https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/ (dostęp: 4.08.2024).

Filipova, I.A. (2024). Legal Regulation of Artificial Intelligence: Experience of China, *Journal of Digital Technologies and Law*, 2(1), s. 46–73.

Flisak, D. (2019). Wpływ rozwoju nowoczesnych technologii na proces stanowienia prawa w Polsce i wybranych państwach Unii Europejskiej, *Zeszyty Prawnicze BAS*, 3(63), s. 194–211.

Flisak, D. (2024). *Akt w sprawie sztucznej inteligencji, komentarz praktyczny*, LEX/el. 2024.

Frazão, A. (2024). Regulation of Artificial Intelligence in Brazil: Examination of Draft Bill no. 2338/2023, *UNIO – EU Law Journal*, 10(1), s. 54–69.

Grupa Ekspertów Wysokiego Szczebla do spraw Sztucznej Inteligencji (2019). *Wytyczne w zakresie etyki dotyczące godnej zaufania sztucznej inteligencji*, https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (dostęp: 1.08.2024).

Habuka, H. (2023). *Japan's Approach to AI Regulation and Its Impact on the 2023 G7 Presidency*, https://www.csis.org/analysis/japans-approach-ai-regulation-and-its-impact-2023-g7-presidency (dostęp: 4.08.2024).

Harasimiuk, D.E., Braun, T. (2021). *Regulating Artificial Intelligence. Binary Ethics and the Law*. London–New York: Routledge.

Hsiung, E., Sung, A. (2024). *Artificial Intelligence 2024. Taiwan*, https://practiceguides.chambers.com/practice-guides/artificial-intelligence-2024/taiwan (dostęp: 2.08.2024).

Hsu, C.W. (2024). The Current Status and Prospects of Artificial Intelligence Regulations in Taiwan, *Journal of AI Law and Regulation*, 1(1), s. 150–154.

IMDA and PDPC (2020). *Model Artificial Intelligence Governance Framework*, https://www.pdpc.gov.sg/help-and-resources/2020/01/model-ai-governance-framework (dostęp: 2.08.2024).

ICO (2024). *UK GDPR Guidance and Resources. Artificial Intelligence*, https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/ (dostęp: 3.08.2024).

Keller, A., Pereira, C.M., Pires, M.L. (2024). The European Union's Approach to Artificial Intelligence and the Challenge of Financial Systemic Risk. W: *Multidisciplinary Perspectives on Artificial Intelligence and the Law* (s. 415–439), H. Sousa Antunes, P.M. Freitas, A.L. Oliveira, C. Martins Pereira, E. Vaz de Sequeira, L. Barreto Xavier (Eds.). Cham: Springer.

Keller, C.I., Magalhães, J.C. (2023). Regulating AI in Democratic Erosion: Context, Imaginaries and Voices in the Brazilian Debate. W: *Elgar Companion to Regulating AI and Big Data in Emerging Economies* (s. 183–200), G.C.K. Yew, M. Yip (Eds.). Northampton: Edward Elgar Publishing.

Khan, M.S., Shoaib, A., Arledge, E. (2024). How to Promote AI in the US Federal Government: Insights from Policy Process Frameworks, *Government Information Quarterly*, 41(1), 101908.

Ko, H.K. (2024). AI Legislation, *Analysis of AI Regulatory Frameworks in South Korea*, https://law.asia/ai-regulatory-frameworks-south-korea/ (dostęp: 30.07.2024).

Ko, H., Park, S. (2020). How to De-Identify Personal Data in South Korea: an Evolutionary Tale, *International Data Privacy Law*, 10(4), s. 385–394.

Komisja Europejska (2020). *Biała księga Komisji w sprawie sztucznej inteligencji. Europejskie podejście do doskonałości i zaufania*. Bruksela, 19.02.2020, COM(2020) 65 final.

Komisja Europejska (2019). *Komunikat Komisji do Parlamentu Europejskiego, Rady, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów. Budowanie zaufania do sztucznej inteligencji ukierunkowanej na człowieka*. Bruksela, 8.04.2019 r., COM(2019) 168 final.

Li, S., Schütte, B., Sankari, S. (2023). The Ongoing AI-Regulation Debate in the EU and Its Influence on the Emerging Economies: A New Case for the 'Brussels Effect'?. W: *Elgar Companion to Regulating AI and Big Data in Emerging Economies* (s. 22–41), M. Findlay, L.M. Ong, W. Zhang (Eds.). Northampton: Edward Elgar Publishing.

Lim, H.Y.F. (2022). Regulatory Compliance. W: *Artificial Intelligence. Law and Regulation* (s. 85–108), C. Kerrigan (Ed.). Northampton: Edward Elgar Publishing.

Mantelero, A. (2022). *Beyond Data. Human Rights, Ethical and Social Impact Assessment in AI*. The Hague: T.M.C. Asser Press.

Martin-Bariteau, F., Scassa, T. (2021). Artificial Intelligence and the Law in Canada. W: *Artificial Intelligence and the Law in Canada*. Toronto: LexisNexis.

Ministry of Foreign Affairs of the People's Republic of China (2022). *Position Paper of the People's Republic of China on Strengthening Ethical Governance of Artificial Intelligence (AI)*, https://www.mfa.gov.cn/mfa_eng/zy/wjzc/ (dostęp: 2.08.2024).

Monetary Authority of Singapore (2024). *Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector*, https://www.mas.gov.sg/ (dostęp: 4.08.2024).

Nowakowski, M. (2023). *Sztuczna inteligencja. Praktyczny przewodnik dla sektora innowacji finansowych*. Warszawa: Wolters Kluwer.

OECD (2019). *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449.

Roberts, H., Babuta, A., Morley, J., Thomas, C., Taddeo, M., Floridi, L. (2023). Artificial Intelligence Regulation in the United Kingdom: A Path to Good Governance and Global Leadership?. *Internet Policy Review*, 12(2).

Roberts, H., Cowls, J., Morley, J., Taddeo, M., Wang, V., Floridi, L. (2021). The Chinese Approach to Artificial Intelligence: An Analysis of Policy, Ethics, and Regulation, *AI and Society*, 36, s. 59–77.

Rolf, S. (2023). *China's Regulations on Algorithms. Context, Impact, and Comparisons with the EU, Friedrich Ebert Stiftung Briefing*, https://futureofwork.fes.de/news-list/e/new-policy-briefing-chinas-regulations-on-algorithms.html (dostęp: 3.08.2024).

Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r. w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46/WE (ogólne rozporządzenie o ochronie danych)

Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r. w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji oraz zmiany rozporządzeń (WE) nr 300/2008, (UE) nr 167/2013, (UE) nr 168/2013, (UE) 2018/858, (UE) 2018/1139 i (UE) 2019/2144 oraz dyrektyw 2014/90/UE, (UE) 2016/797 i (UE) 2020/1828 (akt w sprawie sztucznej inteligencji).

Singh, J.K.S. (2023). The Values of an AI Ethical Framework for a Developing Nation: Considerations for Malaysia. W: *Elgar Companion to Regulating AI and Big Data in Emerging Economies* (s. 115–134), G.C.K. Yew, M. Yip (Eds.). Northampton: Edward Elgar Publishing.

Social Principles of Human-Centric AI (2019). *Japan*, https://oecd.ai/en/dashboards/policy-initiatives/http:%2F%2Faipo.oecd.org%2F2021-data-policyInitiatives-24341 (dostęp: 2.08.2024).

Tabassi, E. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, https://nvlpubs.nist.gov/nistpubs/ai/ (dostęp: 5.08.2024).

The Canadian Labour Congress (2023). *The Artificial Intelligence and Data Act (AIDA)*, Submitted to the Study of Bill C-27. House of Commons' Standing Committee on Industry and Technology (INDU).

UNESCO (2021). *Recommendation on the Ethics of Artificial Intelligence*, SHS/BIO/REC-AIETHICS/2021.

Walter, Y. (2024). Managing the Race to the Moon: Global Policy and Governance in Artificial Intelligence Regulation – A Contemporary Overview and an Analysis of Socioeconomic Consequences, *Discover Artificial Intelligence*, 4(1), s. 14.

Weber, R.H. (2021). Global Law in the Face of Datafication. W: *Artificial Intelligence and International Economic Law*, P. Shin-Yi, L. Ching-Fu, T. Streinz (Eds.). Cambridge: Cambridge University Press.

White House Office of Science and Technology Policy (2022). *The Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People*, https://www.whitehouse.gov/ostp/ai-bill-of-rights (dostęp: 5.08.2024).

12.0

WPŁYW SZTUCZNEJ INTELIGENCJI NA KARIERĘ ZAWODOWĄ W POSTRZEGANIU STUDENTÓW UNIWERSYTETU EKONOMICZNEGO W KRAKOWIE

Kamila Kwiatkowska
Uniwersytet Ekonomiczny w Krakowie

Streszczenie

Celem niniejszego rozdziału jest identyfikacja wpływu sztucznej inteligencji na karierę zawodową w postrzeganiu studentów Uniwersytetu Ekonomicznego w Krakowie. Przeprowadzono badanie ankietowe wśród 132 studentów UEK, jednej z wiodących uczelni ekonomicznych w Polsce. Wyniki pokazują, że studenci dostrzegają zarówno korzyści (np. zwiększenie efektywności pracy, nowe możliwości kariery), jak i wyzwania (np. potencjalna utrata miejsc pracy, obawy o prywatność) związane z rozwojem AI. Większość respondentów (87%) uważa, że AI wpłynie na ich wybór kariery, jednak dla 88% nie zmieniło to ich podejścia do wyboru zawodu. Technologia informacyjna jest postrzegana jako najbardziej obiecujący sektor w kontekście rozwoju AI. Badanie ujawniło również, że 57% studentów potwierdza dostępność programów nauczania związanych z AI na ich uczelni. Rozdział kończy się wnioskami dokumentującymi implikacje dla uczelni wyższych, doradców zawodowych i decydentów politycznych w zakresie kształtowania programów edukacyjnych i strategii przygotowujących studentów do wyzwań rynku pracy w erze sztucznej inteligencji.

Słowa kluczowe: decyzje zawodowe, edukacja wyższa, rynek pracy, sztuczna inteligencja

Wprowadzenie

Rozwój sztucznej inteligencji (AI) w ostatnich latach znacząco zmienił wiele aspektów życia społecznego i gospodarczego, w tym rynek pracy i edukację wyższą. Dynamiczne zmiany technologiczne stawiają przed studentami nowe wyzwania i możliwości, jednocześnie kształtując ich decyzje dotyczące przyszłej kariery zawodowej.

Celem niniejszego rozdziału jest omówienie, w jaki sposób sztuczna inteligencja wpływa na karierę zawodową w opinii studentów Uniwersytetu Ekonomicznego w Krakowie. W rozdziale przedstawiono różne spojrzenia na temat sztucznej inteligencji, omówiono zmiany, które zostały spowodowane wprowadzaniem tych technologii w różnych sektorach, a także odczucia studentów na temat przyszłej kariery zawodowej w dzisiejszym świecie. Kluczowym elementem opracowania są wyniki badania ankietowego, przeprowadzonego wśród studentów Uniwersytetu Ekonomicznego w Krakowie (UEK), jednej z wiodących uczelni ekonomicznych w Polsce.

Analiza ta ma na celu nie tylko przedstawienie obecnej sytuacji, ale także dostarczenie wskazówek dla uczelni wyższych, doradców zawodowych i decydentów politycznych w zakresie kształtowania programów edukacyjnych i strategii przygotowujących studentów do wyzwań rynku pracy w erze sztucznej inteligencji.

12.1. Plany zawodowe studentów w czasach rozwoju sztucznej inteligencji

W literaturze naukowej można zaobserwować zainteresowanie podejmowaniem wyborów zawodowych w czasach rozwoju sztucznej inteligencji i odczuwalnych zmian, które z niej płyną. Lilla Vicsek, Tamás Bokor i Gyöngyvér Pataki [2022, s. 21–35] w latach 2019–2020 zbadali oczekiwania młodych ludzi wobec automatyzacji i przyszłości pracy. W badaniu wzięło udział 62 węgierskich studentów kierunków nietechnicznych, takich jak prawo, rachunkowość, zarządzanie czy marketing. Analiza wywiadów wykazała, że studenci oczekują stopniowych zmian, a nie rewolucji związanych z automatyzacją. Na początku rozmów nie przewidywali oni dużych zmian, ale później, w ich trakcie, zaczęli dostrzegać możliwość większego wpływu technologii na różne zawody. Nie oczekiwali jednak dużego negatywnego wpływu automatyzacji na pracę, przewidując, że dotknie głównie pracowników fizycznych, bez poważnych problemów społecznych. Automatyzacja i AI nie stanowiły kluczowych kwestii w planowaniu ich kariery. Byli przekonani, że wyższe wykształcenie ich ochroni, mimo że nie widzieli potrzeby rozwijania umiejętności technicznych. Badanie sugeruje, że młodzi ludzie mają bardziej optymistyczne oczekiwania wobec przyszłości

pracy niż eksperci, co może prowadzić do braku przygotowania na zmiany na rynku pracy. Autorzy rekomendują dalsze badania i polityki edukacyjne, promujące rozwój umiejętności technicznych.

W 2021 r. Ali Akaner [2021, s. 928] zbadał wpływ sztucznej inteligencji na wybór radiologii jako specjalizacji przez studentów medycyny w Arabii Saudyjskiej. W ramach analizy przeprowadził ankietę wśród 319 studentów z trzech uczelni medycznych w prowincji Al-Qassim. Wyniki pokazały, że 26,96% respondentów uznało radiologię za jedną z trzech najbardziej pożądanych specjalizacji. Mimo że jedynie 23,2% studentów przewidywało możliwość całkowitego zastąpienia radiologów przez AI, 22,26% uczestników badania, którzy nie dostrzegali potencjału AI jako narzędzia wspierającego diagnostykę, w mniejszym stopniu interesowało się wyborem tej ścieżki zawodowej. Studenci przejawiający większe zainteresowanie radiologią oraz posiadający wcześniejsze doświadczenie w obszarze AI wykazywali lepsze zrozumienie tej technologii oraz odczuwali mniejszy poziom niepewności wobec jej zastosowań. Autor sugeruje, że zwiększenie poziomu edukacji dotyczącej sztucznej inteligencji oraz jej rzeczywistego wpływu na praktykę kliniczną może przyczynić się do poprawy percepcji radiologii jako potencjalnej ścieżki kariery zawodowej.

Nurul Sukma Lestari, Dendy Rosman i Trias Septyoari Putranto przeprowadzili w 2021 r. [s. 690–694] badanie, które miało na celu określenie związku między świadomością studentów hotelarstwa na temat robotów, sztucznej inteligencji i automatyzacji usług (RAISA) a ich postrzeganiem przyszłych możliwości kariery zawodowej. W badaniu wzięło udział 195 studentów hotelarstwa z różnych uczelni w Dżakarcie. Dane o charakterze ilościowym zebrano za pomocą kwestionariusza ankiety, a do analizy wykorzystano korelację i regresję wielokrotną. Wyniki badania wykazały pozytywną korelację między świadomością RAISA a postrzeganymi możliwościami kariery oraz między świadomością RAISA a kompetencjami zawodowymi studentów. Zaobserwowano również silną pozytywną korelację między kompetencjami zawodowymi a postrzeganymi możliwościami kariery. Analiza regresji potwierdziła, że świadomość RAISA i kompetencje zawodowe istotnie wpływają na postrzegane możliwości kariery. Badanie wykazało, że studenci hotelarstwa (pokolenie Z) postrzegają RAISA raczej jako szansę niż zagrożenie dla swojej przyszłej kariery. Wyniki sugerują, że programy nauczania powinny uwzględniać nowoczesne technologie, aby lepiej przygotować studentów do przyszłej pracy w branży hotelarskiej.

W 2023 r. Yujie Cao, Azhan Abdul Aziz i Wan Nur Rukiah Mohd Arshard przeprowadzili analizę nastawienia studentów architektury wnętrz do AI i jej potencjalnego wpływu na ich przyszłe kariery. Badanie objęło grupę 230 chińskich studentów trzeciego roku tego kierunku, z czego uwzględniono odpowiedzi 158 osób. Wykorzystano

kwestionariusz bazujący na modelu akceptacji technologii (TAM). Głównym celem było zbadanie postrzegania nowoczesnych narzędzi AI, takich jak ChatGPT, Stable Diffusion czy Midjourney, przez przyszłych projektantów wnętrz oraz ich gotowości do implementacji tych technologii w pracy zawodowej. Wyniki ujawniły, że studenci dysponują ograniczoną wiedzą na temat tych innowacji i wyrażają pewne obawy odnośnie do wpływu AI na rynek pracy. Wykazali oni jednak zainteresowanie wykorzystaniem AI jako wsparcia dla swojej kreatywności i efektywności. Analiza z zastosowaniem modelowania równań strukturalnych potwierdziła, że kluczowymi czynnikami wpływającymi na akceptację AI są jej postrzegana użyteczność i łatwość obsługi. Badacze sugerują, że instytucje edukacyjne powinny położyć większy nacisk na rozwijanie umiejętności technologicznych studentów, aby lepiej przygotować ich do wyzwań zautomatyzowanego rynku pracy w sektorze projektowym.

Badania przeprowadzone przez Collen Flaherty w 2024 r. w ramach serii Student Voice dotyczyło postrzegania sztucznej inteligencji przez studentów w kontekście ich planów akademickich i zawodowych. Wzięło w nich udział 1250 studentów z 49 uczelni w USA. Dane zebrano za pomocą ankiety online. Analiza wyników wykazała, że 14% studentów twierdziło, że AI miała znaczący wpływ na ich plany dotyczące kierunku studiów, a 34% uważało, że wpłynęła w pewnym stopniu. W przypadku planów zawodowych 11% respondentów przyznało, że AI wywarła znaczący wpływ na ich plany, a 31% wskazało, że wpłynęła w pewnym stopniu. Młodsi studenci przejawiali większą świadomość wpływu AI na ich przyszłe kariery niż starsi. Badanie wykazało, że 72% respondentów uważa, że ich uczelnie powinny przygotowywać ich do pracy ze sztuczną inteligencją. Studenci chcą być uczeni etyki korzystania z tej technologii (74%), rozwijać umiejętności krytycznego myślenia (62%) oraz otrzymać praktyczne szkolenie z zastosowania AI (60%). Wyniki wskazują na rosnącą świadomość studentów na temat wpływu sztucznej inteligencji na ich przyszłość edukacyjną i zawodową, oraz oczekiwanie, że uczelnie dostosują się do tych zmian.

12.2. Potrzeba nowych kompetencji pracowniczych i zmian w strukturze zatrudnienia w świetle rozwoju sztucznej inteligencji

W erze sztucznej inteligencji coraz bardziej pożądane stają się umiejętności i kwalifikacje związane z nowymi technologiami. Spektakularny sukces OpenAI wynikał z możliwości prostego korzystania z modelu LLM (Large Language Model). Instrument GPT (Generative Pre-trained Transformer), opracowany przez OpenAI, jest przykładem modelu LLM, który został wytrenowany na ogromnym zbiorze danych tekstowych

w celu przewidywania kolejnych słów w sekwencji [Brown i in., 2023, s. 9]. Raport LinkedIn [2020, s. 7–21] *2020 Emerging Jobs Report* wskazuje, że wśród najszybciej rozwijających się zawodów znajdują się specjaliści ds. AI, inżynierowie robotyki i specjaliści ds. cyberbezpieczeństwa. Z kolei World Economic Forum [2020, s. 7–21] podkreślało znaczenie umiejętności, takich jak krytyczne myślenie, rozwiązywanie złożonych problemów, kreatywność i inteligencja emocjonalna. W świecie kształtowanym przez sztuczną inteligencję i nowe technologie kluczem do sukcesu zawodowego staje się nie tylko specjalistyczna wiedza techniczna, ale także rozwój uniwersalnych kompetencji miękkich. To połączenie umiejętności technologicznych i ludzkich cech będzie definiować przyszłość rynku pracy, stawiając przed nami wyzwanie ciągłego rozwoju i adaptacji do zmieniającej się rzeczywistości zawodowej. Najnowszy raport World Economic Forum [2023, s. 6] jako kluczowe umiejętności wciąż podaje przede wszystkim myślenie analityczne i kreatywne.

Technologia sztucznej inteligencji zmienia zapotrzebowanie na różne zawody i umiejętności. Zgodnie z raportem McKinsey Global Institute [2017, s. 8] do 2030 r. około 375 mln pracowników (14% globalnej siły roboczej) może potrzebować zmiany kategorii zawodowej. Automatyzacja ma znaczący wpływ na tradycyjne miejsca pracy.

Według badania PwC [2018, s. 7] do połowy 2030 r. około 30% miejsc pracy może być zagrożonych usunięciem. Choć niektóre zawody znikną, pojawią się nowe możliwości, wymagające od pracowników elastyczności i gotowości do ciągłego rozwoju.

World Economic Forum w *The Future of Jobs Report 2020* [2020, s. 5] podało, że do 2025 r. 85 mln miejsc pracy może zostać zastąpionych przez maszyny, ale jednocześnie może powstać 97 mln nowych stanowisk, lepiej dostosowanych do nowego podziału pracy między ludźmi, maszynami i algorytmami.

W najnowszym raporcie World Economic Forum [2023, s. 7] możemy przeczytać, że organizacje realizują 34% zadań z użyciem maszyn, a pozostałe 66% wykonują ludzie. Oznacza to nieznaczny wzrost automatyzacji o 1% w porównaniu z szacunkami respondentów z edycji raportu z 2020 r.

Na potrzeby przygotowania niniejszego rozdziału sprawdzono dostępność publikacji zawierających słowa kluczowe: „artificial intelligence", „students", „career", „decisions". Wyszukiwanie w bazie Scopus dało wynik 125 publikacji w języku angielskim. Publikacje dotyczą głównie wykorzystania sztucznej inteligencji, uczenia maszynowego i innych zaawansowanych technologii w edukacji, szczególnie w kontekście wspomagania wyboru kariery i kierunku studiów przez studentów, przewidywania ich sukcesu edukacyjnego i zawodowego, personalizacji procesu nauczania oraz analizy danych edukacyjnych w celu poprawy jakości kształcenia. Najczęściej pojawiające się kierunki studiów to informatyka i nauki komputerowe, inżynieria (różne

specjalności), medycyna i nauki o zdrowiu (zwłaszcza radiologia), biznes i zarządzanie oraz nauki ścisłe, technologia, inżynieria i matematyka. Badania koncentrują się głównie na studentach szkół wyższych, uczniach szkół średnich rozważających wybór kierunku studiów oraz absolwentach wchodzących na rynek pracy. Stosowane są różnorodne metody badawcze, w tym analiza danych z wykorzystaniem uczenia maszynowego, badania ankietowe i wywiady, projektowanie i testowanie systemów rekomendacyjnych oraz analizy długoterminowych trendów edukacyjnych i zawodowych. Najwięcej publikacji pochodzi z ostatnich kilku lat (2020–2023), co wskazuje na rosnące zainteresowanie tą tematyką w kontekście szybkiego rozwoju AI i innych technologii. Większość opisanych badań pochodzi zaś z USA lub Azji. Istnieje zatem odczuwalna potrzeba ustalenia, jak narzędzia AI mogą wpływać na karierę zawodową młodych ludzi.

12.3. Metodyka badań

Na potrzeby niniejszego rozdziału przeprowadzono badania ilościowe z wykorzystaniem ankiety na platformie Google Forms. Kwestionariusz składał się głównie z pytań zamkniętych, ale umożliwiał też wpisanie własnych odpowiedzi. Dane z ankiety analizowano przy użyciu arkusza kalkulacyjnego Microsoft Excel.

Ankieta była anonimowa, a uczestnicy zostali poinformowani, że wyniki mogą zostać użyte do celów naukowych. Pojawiło się w niej osiem pytań dotyczących przyszłej ścieżki zawodowej, przewidywanych korzyści i wyzwań związanych z rozwojem sztucznej inteligencji, obiecujących sektorów w kontekście rozwoju tych technologii, dostępności programów nauczania związanych z AI na uczelni oraz przygotowania do powiązanych z tym wyzwań.

Badania przeprowadzono na przełomie kwietnia i maja 2024 r. wśród studentów Uniwersytetu Ekonomicznego w Krakowie. Wybór tej grupy badawczej pozwala na uzyskanie cennych informacji o perspektywach i obawach przyszłych specjalistów w dziedzinach związanych z ekonomią, zarządzaniem i finansami w kontekście rosnącej roli nowoczesnych technologii, z naciskiem na sztuczną inteligencję. W badaniu uczestniczyło 132 respondentów, a wszystkie odpowiedzi zostały uwzględnione.

Próba badawcza składała się z 132 osób, 70% stanowiły kobiety, a 30% mężczyźni. Analiza wieku respondentów wykazała, że 96% z nich miało 18–26 lat, a 4% znalazło się w przedziale wiekowym 27–35 lat. Ponadto 26% respondentów mieszkało na wsi, 20% w miastach do 50 tys. mieszkańców, 8% pochodziło z miast od 150 do 500 tys. mieszkańców, a 46% z miast powyżej 500 tys. mieszkańców. Wśród ankietowanych

28% to studenci studiów niestacjonarnych, pozostali studiują stacjonarnie. Najczęściej wskazywane kierunki studiów to: analityka gospodarcza, zarządzanie, finanse i rachunkowość, rachunkowość i controlling oraz prawo.

12.4. Wyniki badań

Badania pokazują, że zdecydowana większość studentów dostrzega pewne zależności między rozwojem sztucznej inteligencji a swoim przyszłym życiem zawodowym. Zgodnie z wynikami przedstawionymi w tabeli 12.1 aż 87% respondentów uważa, że sztuczna inteligencja wpłynie na ich wybór kariery, z czego 57% jest o tym zdecydowanie przekonanych, a 30% raczej się z tym zgadza. Tylko niewielki odsetek badanych (13%) nie dostrzega takiego wpływu, z czego zaledwie 3% zdecydowanie odrzuca taką możliwość.

Tabela 12.1. Rozkład odpowiedzi na pytanie: „Czy uważasz, że sztuczna inteligencja wpłynie na wybór Twojej przyszłej ścieżki zawodowej?" (%)

Odpowiedzi	Odsetek odpowiedzi
Zdecydowanie tak	57
Raczej tak	30
Raczej nie	10
Zdecydowanie nie	3

Źródło: opracowanie własne.

Tabela 12.2. Rozkład odpowiedzi na pytanie: „Jakie korzyści dostrzegasz w związku z wykorzystaniem sztucznej inteligencji w przyszłej pracy?" (%)

Odpowiedzi	Odsetek odpowiedzi
Zwiększenie efektywności i automatyzacja procesów pracy	86
Nowe możliwości kariery w obszarach związanych z AI	43
Poprawa jakości usług i produktów	36
Innowacyjne rozwiązania w branży	47
Nie dostrzegam żadnych korzyści	1

Źródło: opracowanie własne.

Studenci dostrzegają liczne korzyści związane z wykorzystaniem AI w przyszłej pracy (tabela 12.2). Zdecydowana większość (86%) jako główną zaletę wskazuje zwiększenie efektywności i automatyzację procesów pracy. Prawie połowa respondentów (47%) dostrzega potencjał AI w tworzeniu innowacyjnych rozwiązań w branży, a 43% widzi nowe możliwości kariery w obszarach związanych ze sztuczną inteligencją. Poprawę jakości usług i produktów zauważa 36% badanych. Co ciekawe, tylko 1% respondentów nie wymienia żadnych korzyści, co świadczy o powszechnym przekonaniu o pozytywnym wpływie AI na przyszłe środowisko pracy.

Mimo wskazywania licznych zalet AI studenci mają również obawy związane z wykorzystywaniem jej w przyszłej pracy (tabela 12.3). Największe z nich to potencjalna utrata miejsc pracy, wyrażona przez 73% badanych, oraz obawy dotyczące prywatności i bezpieczeństwa danych (64%). Konieczność ciągłego doskonalenia umiejętności stanowi wyzwanie dla 43% respondentów, a etyczne dylematy związane z autonomią maszyn niepokoją 23% badanych. Warto zauważyć, że żaden z respondentów nie wskazał braku wyzwań, co świadczy o świadomości złożoności problemu wśród studentów.

Tabela 12.3. Rozkład odpowiedzi na pytanie: „Jakie wyzwania dostrzegasz w związku z wykorzystaniem sztucznej inteligencji w przyszłej pracy?" (%)

Odpowiedzi	Odsetek odpowiedzi
Potencjalna utrata miejsc pracy	73
Obawy dotyczące prywatności i bezpieczeństwa danych	64
Konieczność ciągłego doskonalenia umiejętności	43
Etyczne dylematy związane z autonomią maszyn	23
Nie dostrzegam żadnych wyzwań	0

Źródło: opracowanie własne.

Interesujące wyniki przyniosło badanie oceny wpływu sztucznej inteligencji na atrakcyjność poszczególnych zawodów. Zdecydowana większość respondentów (88%) deklaruje, że rozwój AI nie zmienił ich podejścia do wyboru zawodu. Jednak wśród pozostałych 12% badanych zauważono zwiększone zainteresowanie karierami w takich obszarach, jak: programowanie, analiza danych, informatyka i grafika komputerowa, z równym udziałem 3% dla każdej z tych specjalizacji.

Studenci mają wyraźnie wyklarowane opinie na temat sektorów, które uważają za najbardziej obiecujące z perspektywy rozwoju AI (tabela 12.4). Pierwsze miejsce zajęła technologia informacyjna, wskazana przez 77% respondentów. Na drugim miejscu ex aequo znalazły się zdrowie i medycyna oraz finanse i bankowość (po 39%). Inżynie-

ria i przemysł są postrzegane jako obiecujące przez 28% badanych, a edukacja przez 36%. Transport i logistyka zamykają listę z 27% wskazań.

Badanie wykazało również, że uczelnie zaczynają dostrzegać potrzebę kształcenia w zakresie sztucznej inteligencji. Aż 57% respondentów potwierdza, że ich uczelnia oferuje programy nauczania związane ze sztuczną inteligencją, podczas gdy 43% wskazuje na brak takich programów.

Tabela 12.4. Rozkład odpowiedzi na pytanie: „Jakie sektory lub dziedziny uważasz za najbardziej obiecujące z perspektywy rozwoju sztucznej inteligencji?" (%)

Odpowiedzi	Odsetek odpowiedzi
Technologia informacyjna	77
Zdrowie i medycyna	39
Finanse i bankowość	39
Edukacja	36
Inżynieria i przemysł	28
Transport i logistyka	27

Źródło: opracowanie własne.

Podsumowanie

Zdecydowana większość badanych studentów (87%) uważa, że AI wpłynie na ich wybór kariery. Wynik ten jest znacząco wyższy niż w badaniu Flaherty [2024], podczas którego 47% amerykańskich studentów dostrzegało wpływ AI na ich plany zawodowe. Może to sugerować wyższą świadomość polskich studentów w zakresie potencjalnego oddziaływania AI na rynek pracy lub większe obawy związane z tym zjawiskiem. Jednocześnie, podobnie jak w badaniu Vicsek i in. [2022], studenci wydają się postrzegać zmiany związane z AI jako stopniowe, a nie rewolucyjne. Świadczy o tym fakt, że aż 88% respondentów deklaruje, iż rozwój AI nie zmienił ich podejścia do wyboru zawodu. To optymistyczne podejście może jednak prowadzić do niedostatecznego przygotowania na nadchodzące zmiany na rynku pracy, co sugerują również Vicsek i in. [2022]. Ankietowani studenci dostrzegają zarówno korzyści, jak i wyzwania związane z rozwojem AI. Główną zaletą, wskazywaną przez 86% badanych, jest zwiększenie efektywności i automatyzacja procesów pracy. Wynik ten koresponduje z ogólnym trendem postrzegania AI jako narzędzia zwiększającego produktywność [World Economic Forum, 2023]. Ankietowani wyrażają obawy, dotyczące głównie potencjalnej

utraty miejsc pracy (73%) oraz prywatności i bezpieczeństwa danych (64%). Te obawy są zbieżne z wynikami innych badań, np. raportu McKinsey Global Institute [2017], który przewiduje, że do 2030 r. około 14% globalnej siły roboczej może potrzebować zmiany kategorii zawodowej ze względu na automatyzację.

Wyniki badania sugerują potrzebę dalszego dostosowywania programów nauczania do wyzwań ery AI. Potwierdziły one przypuszczenia autorki dotyczące dostrzegania zmian przez grupę badanych studentów, wynikających z analizy podobnych badań na ten temat. Uczelnie powinny nie tylko oferować kursy techniczne związane z AI, ale także kłaść nacisk na rozwój umiejętności miękkich, takich jak krytyczne myślenie i kreatywność, które – według World Economic Forum [2023] – pozostają kluczowe na rynku pracy.

Ponadto, biorąc pod uwagę obawy studentów dotyczące utraty miejsc pracy, istotne jest, aby programy edukacyjne uwzględniały również aspekty etyczne i społeczne związane z rozwojem AI, co pomoże studentom lepiej zrozumieć i przygotować się na nadchodzące zmiany.

Należy zauważyć, że badanie zostało przeprowadzone na jednej uczelni ekonomicznej, na niewielkiej liczbie studentów, co może ograniczać możliwość generalizacji wyników na całą populację studentów w Polsce. Przyszłe badania mogłyby objąć szerszy zakres uczelni ekonomicznych i kierunków studiów, aby uzyskać bardziej kompleksowy obraz sytuacji.

Bibliografia

Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A. Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D.M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., Amodei, D. (2020). *Language Models Are Few-Shot Learners*. DOI: 10.48550/arXiv.2005.14165.

Cao, Y., Aziz, A.A., Arshard, W.N.R.M. (2023). University Students' Perspectives on Artificial Intelligence: A Survey of Attitudes and Awareness Among Interior Architecture Students, *International Journal of Educational Research and Innovation*, 20, s. 1–21.

Flaherty, C. (2024). *Survey: How AI Is Impacting Students' Career Choices*, https://www.insidehighered.com/news/student-success/life-after-college/2024/01/10/survey-college-students-thoughts-ai-and-careers (dostęp: 16.08.2024).

Lestari N.S., Rosman, D., Putranto, T.S. (2021). The Relationship Between Robot, Artificial Intelligence, and Service Automation (RAISA) Awareness, Career Competency, and Perceived Career Opportunities: Hospitality Student Perspective. W: *2021 International Conference*

on Information Management and Technology (s. 690–695). Jakarta, Indonesia. DOI: 10.1109/ICIMTech53080.2021.9535054.

LinkedIn (2020). *2020 Emerging Jobs Report*, https://business.linkedin.com/content/dam/me/business/en-us/talent-solutions/emerging-jobs-report/Emerging_Jobs_Report_U.S._FINAL.pdf (dostęp: 27.07.2024).

McKinsey Global Institute (2017). *Jobs Lost, Jobs Gained: Workforce Transitions in a Time of Automation*, https://www.mckinsey.com/featured-insights/future-of-work/jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and-wages (dostęp: 27.07.2024).

PwC (2018). *Will Robots Really Steal Our Jobs? An International Analysis of The Potential Long Term Impact of Automation*, https://www.pwc.com/h/hu/kiadvanyok/assets/pdf/impact_of_automation_on_jobs.pdf (dostęp: 27.07.2024).

Vicsek, L., Bokor, T., Pataki, G.F. (2022). Younger Generations' Expectations Regarding Artificial Intelligence in the Job Market: Mapping Accounts About the Future Relationship of Automation and Work, *Journal of Sociology*, 60, s. 21–38.

World Economic Forum (2020). *The Future of Jobs Report 2020*, https://www.weforum.org/reports/the-future-of-jobs-report-2020 (dostęp: 27.07.2024).

World Economic Forum (2023). *The Future of Jobs Report 2023*, https://www.weforum.org/reports/the-future-of-jobs-report-2023 (dostęp: 7.08.2024).

IV

SEKTOROWE ZASTOSOWANIA SZTUCZNEJ INTELIGENCJI

13.0

ZASTOSOWANIE SZTUCZNEJ INTELIGENCJI I HIPERAUTOMATYZACJI W PROCESIE FUZJI PRZEJĘĆ

Paulina Roszkowska
Szkoła Główna Handlowa w Warszawie / Bayes Business School, Londyn

Streszczenie

Sztuczna inteligencja i hiperautomatyzacja stają się częścią transformacji cyfrowej przedsiębiorstw, ich zarządzania strategicznego oraz finansów. Rozdział ten analizuje, w jaki sposób nowoczesne technologie mogą optymalizować procesy fuzji i przejęć, usprawniając analizę danych oraz automatyzując powtarzalne zadania, szczególnie na etapach poszukiwania spółki do przejęcia, due diligence, wyceny i integracji po transakcji. Przedstawione przykłady niepowodzeń transakcji, takich jak SoFi – Technisys czy HP – Autonomy, pokazują potencjał technologii w usprawnianiu tych złożonych procesów. Wnioski wskazują, że choć AI i hiperautomatyzacja mogą zwiększyć efektywność i dokładność analiz, nie zastępują one w pełni ludzkiego osądu, a ich skuteczność zależy od odpowiedniego monitorowania i wdrażania w połączeniu z wiedzą ekspercką.

Słowa kluczowe: analiza danych, analiza predykcyjna, fuzje i przejęcia, hiperautomatyzacja, sztuczna inteligencja, zrobotyzowana automatyzacja procesów

[1] Chciałabym podziękować moim studentom: Jennie Carlsson Backlund, Jorge Valero Hay, Tanapat Sirilon i Michele Slomp, uczestniczącym w zajęciach Fintech w różnych szkołach biznesu, za nasze dyskusje oraz ich spostrzeżenia dotyczące przypadków fuzji i przejęć opisanych w tym rozdziale.

Wprowadzenie

Sztuczna inteligencja i hiperautomatyzacja stają się kluczowymi narzędziami w wielu obszarach biznesu, w tym w finansach. Fuzje i przejęcia (*mergers and acquisitions*, M&A) od dawna są jednym z najbardziej dyskutowanych tematów w finansach, pojawiając się na pierwszych stronach najważniejszych gazet biznesowych, w mediach społecznościowych i nie tylko. Fuzje i przejęcia odgrywają istotną rolę w strategii korporacyjnej i rozwoju, dając firmom możliwość osiągnięcia efektów skali, zaistnienia na nowych rynkach oraz zdobycia przewagi konkurencyjnej dzięki synergii [Patel, 2022]. Transakcje M&A w dużej mierze opierają się na czynniku ludzkim, są specyficzne, dlatego nie mogą być w pełni zautomatyzowane. Niemniej coraz więcej elementów procesu M&A jest wspomaganych przez sztuczną inteligencję i hiperautomatyzację, co może wynikać z licznych przykładów nieudanych transakcji.

W niniejszym rozdziale przeanalizowane zostanie, które części procesu M&A mogą zostać zautomatyzowane w celu optymalizacji wyników fuzji i przejęć, w szczególności uniknięcia transakcji obniżających wartość. Zaprezentowane zostaną również narzędzia automatyzacji, które mogą być wykorzystywane przez spółki przejmujące jako część transformacji cyfrowej w obszarze finansów i strategii, a także przez firmy doradcze, które stanowią nieodłączną część procesu M&A. One również przechodzą transformację cyfrową, co może mieć konsekwencje kosztowe i jakościowe dla spółek przejmujących i przejmowanych.

Typowy proces M&A obejmuje następujące etapy (perspektywa spółki przejmującej) [Moeller, Brady, 2014]:

1. Rozwój strategii korporacyjnej i due diligence.
2. Organizacja procesu fuzji, np. wybór lidera projektu, zespołów, identyfikacja doradców.
3. Wycena transakcji i negocjacje.
4. Strukturyzacja i zatwierdzenie umowy.
5. Integracja po transakcji.
6. Przegląd po przejęciu.

Due diligence, wycena i integracja po fuzji to najbardziej złożone etapy procesu M&A, na których często dochodzi do niepowodzeń transakcji. Dlatego w kolejnych częściach rozdziału zostaną one dokładnie przeanalizowane pod kątem możliwości wykorzystania nowoczesnych technologii w celu ich usprawnienia i ograniczenia ryzyka niepowodzeń. Dodatkowo w analizie uwzględniony zostanie etap poszukiwania spółki przejmowanej (*target company*), ponieważ wybór niewłaściwego celu lub brak świadomości lepszych opcji często prowadzi do nieudanych transakcji. Wybór nie-

właściwego celu w transakcjach M&A rzadko jest bezpośrednio identyfikowany jako główna przyczyna niepowodzenia, ponieważ problemy wynikające z błędnych decyzji na tym etapie często stają się widoczne dopiero po pewnym czasie, gdy pojawiają się trudności z integracją przejętej firmy, konflikty kulturowe lub brak osiągnięcia zakładanych synergii. Wówczas pierwotny wybór celu bywa trudny do jednoznacznego wskazania jako przyczyna problemów. Dodatkowo presja na szybkie osiągnięcie wzrostu za pomocą M&A, zdobycie nowych rynków czy przejęcie technologii często prowadzą do wyboru suboptymalnych spółek do przejęcia, ale ponieważ decyzje te są zgodne z bieżącą strategią, rzadko bywają postrzegane jako główne źródło niepowodzenia transakcji.

W ramach analizy transformacji cyfrowej – i pytania, czy każdy etap procesu M&A może zostać zautomatyzowany – omówione zostaną przykłady najnowszych transakcji, które napotkały różne problemy. Zostanie również pokazane, w jaki sposób nowoczesne technologie mogą rozwiązać te trudności, zwiększając efektywność procesów i minimalizując błędy strategiczne. Dwa rozwiązania technologiczne szczególnie pomocne w fuzjach i przejęciach to:

- Sztuczna inteligencja (*artificial intelligence*, AI) i analityka danych: narzędzia analityki danych wspomagane przez AI wykorzystują zaawansowane rozwiązania technologiczne do bardziej efektywnego czyszczenia, analizy i modelowania danych finansowych i operacyjnych, w celu wspierania procesów podejmowania decyzji opartych na danych [Bakumenko, Elragal, 2022; Krohn, 2023]. W szczególności:
  - narzędzia analizy predykcyjnej, wykorzystujące modele do analizy danych finansowych i rynkowych, pomagają prognozować przyszłą wydajność i identyfikować potencjalne ryzyka; zaawansowane algorytmy AI pozwalają badać historyczny wzrost przychodów, koszty pozyskania klientów, ale także trendy rynkowe i inne bardziej nietypowe dane w celu predykcji przyszłych przychodów i innych parametrów firmy przejmowanej; modele oparte na sztucznej inteligencji (uczenie maszynowe, sieci neuronowe, itd.) są w stanie przetwarzać duże zbiory danych – w ten sposób pomagają identyfikować trendy i anomalie, np. nietypowe transakcje finansowe, które nie są wykrywalne przez tradycyjne metody analizy danych;
  - analiza danych w czasie rzeczywistym, która ułatwia szybkie podejmowanie decyzji, np. w dynamicznych procesach *due dilligence* i wyceny M&A.
- Hiperautomatyzacja oparta na AI i zrobotyzowanej automatyzacji procesów (*robotic process automation*, RPA): RPA to technologia wykorzystująca roboty programowe lub „boty" do automatyzacji powtarzalnych i opartych na regułach zadań. Hiperautomatyzacja rozszerza możliwości RPA poza automatyzację predefiniowanych

zadań, wykorzystując AI do określania najlepszej strategii wykonywania tych działań. W procesie fuzji i przejęć hiperautomatyzacja może usprawnić zadania takie jak [Dilmegani, 2023]:

- analiza dokumentów: hiperautomatyzacja pomaga szybko pobierać i przeglądać dużą liczbę dokumentów, unikając błędów ludzkich i skracając czas poświęcony na te czynności. Obejmuje to m.in. przegląd sprawozdań finansowych, umów, dokumentów prawnych i innych strategicznych informacji w celu wyodrębnienia konkretnych danych, które można następnie poddać dalszej analizie;
- ekstrakcja i analiza danych: hiperautomatyzacja pomaga wyodrębniać dane z wielu źródeł, weryfikować je według zdefiniowanych kryteriów i wprowadzać do oprogramowania do due diligence lub arkuszy kalkulacyjnych procesu wyceny; minimalizuje to ryzyko błędów przy ręcznym wprowadzaniu danych i oszczędza czas;
- modelowanie finansowe: hiperautomatyzacja umożliwia zbieranie danych, aby automatycznie wykonywać obliczenia i tworzyć modele finansowe dotyczące wartości i potencjału firmy przejmowanej;
- automatyzacja przepływu zadań: hiperautomatyzacja pomaga automatyzować procesy przepływu pracy, przypisując zadania odpowiednim członkom zespołu, śledząc postęp i zapewniając dotrzymywanie terminów. Może również wysyłać powiadomienia i przypomnienia członkom zespołu, poprawiając współpracę i efektywność.

W rozdziale pojawiają się także odwołania do innych technologii i konkretnych narzędzi dostępnych na rynku, jeśli te mają szczególne zastosowanie w usprawnianiu procesu M&A.

13.1. Wykorzystanie danych w identyfikacji spółek do przejęcia

Identyfikcja spółek do przejęcia to niezwykle istotny etap M&A, który pozwala wyłonić podmioty o potencjale strategicznym dla spółki przejmującej. Tradycyjne metody, opierające się na osobistych sieciach i ręcznym poszukiwaniu, są czasochłonne i ograniczają zasięg poszukiwań [Ziuznys, 2022]. Metody te mogą również wprowadzać subiektywizm. Nowoczesne rozwiązania technologiczne oparte na AI i analizie danych rewolucjonizują proces poszukiwania spółek do przejęć, czyniąc go bardziej dostępnym, efektywnym i obiektywnym [Clarke, Uhlaner, Wol, 2020].

Przykładem jest platforma DealCloud, integrująca innowacje technologiczne w pozyskiwaniu transakcji dla firm *private equity*, banków inwestycyjnych i zespołów ds. rozwoju korporacyjnego. DealCloud wykorzystuje uczenie maszynowe i analitykę big data do przeszukiwania danych rynkowych, zarządzania relacjami i śledzenia potencjalnych transakcji. Automatyzuje zbieranie danych, dostarcza prognozy i identyfikuje możliwości zgodne z celami strategicznymi firmy. Użytkownicy mogą analizować trendy rynkowe i ruchy konkurencji, dokonywać kompleksowego przeglądu rynku.

DealCloud umożliwia firmom przejście od tradycyjnych metod do bardziej dynamicznego, opartego na danych podejścia, przyspieszając identyfikację wartościowych spółek do przejęcia. Technologia ta pokazuje praktyczne korzyści innowacji technologicznych we współczesnym pozyskiwaniu transakcji.

W lutym 2022 r. SoFi Technologies, amerykańska firma finansowa, przejęła firmę Technisys, specjalistę od bankowości cyfrowej, za około 1,1 mld USD. Celem było rozszerzenie rynku cyfrowych usług finansowych SoFi oraz realizacja wizji firmy jako kompleksowego centrum finansowego. Platforma Cyberbank firmy Technisys miała zwiększyć zdolność SoFi do dostarczania innowacyjnych doświadczeń w bankowości cyfrowej i rozwijania nowych produktów finansowych. Przejęcie poszerzyło ofertę SoFi i wzmocniło jej pozycję w sektorze fintech [SoFi Technologies, 2022].

Przejęcie Technisys przez SoFi pokazuje, jak sztuczna inteligencja i hiperautomatyzacja wpływają na pozyskiwanie transakcji. SoFi wykorzystało analitykę danych i narzędzia AI do oceny pozycji rynkowej, kondycji finansowej i zgodności Technisys z jej celami strategicznymi, co przyspieszyło proces wyboru spółki do przejęcia i analizy due diligence.

Platformy technologiczne, wykorzystujące zaawansowaną analitykę big data, chmurę obliczeniową oraz uczenie maszynowe, zapewniają bardziej efektywne podejście do pozyskiwania transakcji [SourceScrub, 2024]. Hiperautomatyzacja i sztuczna inteligencja umożliwiają głębszą analizę oraz szybsze przetwarzanie danych, przewyższając tradycyjne metody. Technologiczne narzędzia usprawniają pozyskiwanie transakcji poprzez szybkie identyfikowanie potencjalnych celów i oszczędzanie czasu dzięki analizie danych. W przypadku SoFi narzędzia te pomogły w strategicznym przejęciu, umożliwiając dokładną ocenę Technisys.

Integracja technologii na tym etapie fuzji i przejęć wiąże się jednak z ryzykiem. Choć ludzka analiza często prowadzi do pominięcia (*bias*) jakiegoś obszaru (czyt. subiektywizmu), poleganie wyłącznie na ocenach ilościowych opartych na dużych zbiorach (i kategoriach) danych może prowadzić do pominięcia ważnych czynników jakościowych. Należy również pamiętać o kwestiach bezpieczeństwa danych. Firmy

powinny wdrożyć zaawansowane środki cyberbezpieczeństwa i uważnie nadzorować algorytmy AI, aby minimalizować to ryzyko.

Dla wielu firm, rozwijających się poprzez fuzje i przejęcia, zwiększenie efektywności i bezpieczeństwa procesu wstępnego przeglądu potencjalnych spółek do przejęcia poprzez integrację platform technologicznych ma fundamentalne znaczenie. Hiperautomatyzacja powinna być jednak uzupełniana ludzką ekspertyzą, aby zapewnić równowagę pomiędzy rekomendacjami pochodzącymi z analizy danych z ekspercką intuicją i zrozumieniem branży oraz dynamiki procesu M&A. Hybrydowe podejście, łączące technologię z ludzką ekspertyzą, pomaga skutecznie identyfikować i oceniać potencjalne spółki do przejęcia, zwiększając szanse na sukces M&A i wzrost wartości firmy.

13.2. AI, analiza danych i RPA w badaniu due diligence

Due diligence to etap procesu M&A, który dostarcza informacji o wartości i ryzyku spółki przejmowanej, mając na celu ograniczenie zidentyfikowanych zagrożeń, zmniejszenie asymetrii informacji oraz poprawę pozycji negocjacyjnej spółki przejmującej. W przypadku tak dużych transakcji procedura due diligence obejmuje szczegółową, ręczną analizę dokumentów finansowych, która jest wyjątkowo czasochłonna i kosztowna [Hanelt, Firk, Hildebrandt, Kolbe, 2021]. Istnieje wiele czynników, które spowalniają proces due diligence i sprawiają, że jest on bardzo nieefektywny. Aby usprawnić ten proces, konieczne jest nowoczesne podejście do due diligence, obejmujące wykorzystanie zaawansowanych technologii, takich jak AI, analiza danych i automatyzacja procesów.

Narzędzia analityki danych oparte na sztucznej inteligencji oferują predykcje, analizując dane historyczne, aby prognozować wyniki i identyfikować ryzyka. Dzięki analizie w czasie rzeczywistym możliwe jest szybsze podejmowanie decyzji, co odgrywa kluczową rolę w procesach M&A. AI wykrywa anomalie w sprawozdaniach finansowych i odchylenia od standardów branżowych, a zrobotyzowana automatyzacja procesów usprawnia zadania, takie jak ekstrakcja danych i modelowanie finansowe, minimalizując błędy ludzkie i przyspieszając due diligence. Automatyzacja procesów za pomocą RPA może usprawnić przydzielanie zadań i zarządzanie terminami, potencjalnie poprawiając współpracę oraz wydajność zespołów zaangażowanych w transakcję.

W 2011 r. kalifornijska firma informatyczna Hewlett-Packard (HP) przejęła brytyjską firmę oprogramowania Autonomy za 11 mld USD. Jednak już rok później okazało się, że transakcja była porażką, gdy HP musiało zgłosić odpis aktualizujący wartość aktywów w wysokości 8,8 mld USD. Niepowodzenie przypisano zarzutom, że firma

Autonomy sztucznie zawyżyła swoje wyniki finansowe przed przejęciem. Tworzyło to mylący obraz kondycji finansowej firmy i doprowadziło do jej przewartościowania [Bobelian, 2012]. Te rzekome nieprawidłowości obejmowały niewłaściwe rozpoznawanie przychodów oraz błędną klasyfikację różnych kosztów, sprawiając wrażenie wyższej rentowności Autonomy. Stało się jednak jasne, że rozbieżności księgowe nie zostały wykryte na etapie due diligence, sugerując, że HP mogło nie przeprowadzić szczegółowego badania dokumentacji finansowej Autonomy przed przejęciem. Brak tego badania ostatecznie doprowadził do poważnych strat finansowych i sporów prawnych po fuzji.

Gdyby HP zastosowało te technologie podczas przejęcia Autonomy, mogłoby mieć większe szanse na bardziej szczegółowe zbadanie kondycji finansowej oraz zobowiązań kontraktowych firmy. Narzędzia do analityki predykcyjnej mogłyby wspomóc HP w identyfikacji potencjalnych nieprawidłowości w raportach finansowych Autonomy oraz usprawnić monitorowanie ryzyk. Algorytmy AI pomogłyby analizować dokumenty prawne w czasie rzeczywistym, pozwalając na wcześniejsze zidentyfikowanie potencjalnych problemów finansowych, a RPA mogłaby przyspieszyć gromadzenie danych finansowych, redukując ryzyko błędów ludzkich.

W przypadku przejęcia firmy Autonomy i wielu innych transakcji zastosowanie AI i RPA jest szczególnie istotne ze względu na złożoność firmy przejmowanej. Technologie te usprawniłyby analizę dokumentów, identyfikując obszary problemowe, które mogłyby zostać przeoczone podczas tradycyjnej weryfikacji. Dzięki hiperautomatyzacji procesy due diligence stają się szybsze, tańsze i dokładniejsze, umożliwiając lepszą alokację zasobów ludzkich na bardziej strategiczne zadania.

Integracja AI, analityki danych i RPA stanowi krok naprzód w technologii finansowej, szczególnie w kontekście dużych transakcji, jak przejęcie Autonomy. Automatyzacja analizy sprawozdań finansowych i umów przyspiesza identyfikację ryzyk, dostarczając decydentom aktualnych informacji wspierających decyzje strategiczne.

Choć analiza predykcyjna, uczenie maszynowe i RPA oferują liczne korzyści, istnieją także wyzwania, takie jak zależność wyników od jakości i precyzji danych. Błędne, niekompletne dane mogą prowadzić do problemów z dokładnością i wiarygodnością uzyskanych wniosków oraz niewłaściwych decyzji. Dodatkowo, jeśli dane wykorzystywane do oszacowania modelu były stronnicze, algorytmy AI mogą dziedziczyć tę stronniczość, wpływając tym samym na rzetelność wyników. Związane z tym są również kwestie etyczne, prywatności i bezpieczeństwa danych finansowych. Wykorzystanie zewnętrznych narzędzi technologicznych oraz wymiana dużej ilości danych między systemami, które mogą mieć różne protokoły bezpieczeństwa, zwiększają podatność na cyberataki (naruszenie danych i nieautoryzowany dostęp do poufnych informacji)

[Deloitte, 2021], co może prowadzić do szkód finansowych i reputacyjnych, kosztów materialnych i niematerialnych, związanych z odtworzeniem infrastruktury. Dlatego dane należy odpowiednio zabezpieczyć, aby zminimalizować wspomniane ryzyka.

Technologie, o których mowa, powinny służyć wsparciu ludzkich decyzji, a nie ich zastępowaniu. Choć hieperautomatyzacja upraszcza procesy, nadzór ludzki pozostaje niezbędny do uwzględniania subtelnych kwestii, które mogą zostać pominięte przez algorytmy, i zapewnia kompleksowe zrozumienie wyników due diligence [Müller, Päuser, Kurzrock, 2021]. Wzajemne uzupełnianie się technologii i ludzkiej oceny pozwala na bardziej kompleksowe i efektywne przeprowadzanie due diligence w przyszłych transakcjach.

13.3. AI i analiza predykcyjna w procesie wyceny M&A

Wycena firm służy do określenia ich aktualnej wartości, szczególnie istotnej w transakcjach M&A, gdzie proces ten bywa trudny i zależny od przyjętych założeń. Szczególnym wyzwaniem jest oszacowanie wartości aktywów niematerialnych, takich jak technologie, dane czy własność intelektualna, których tradycyjne metody nie potrafią w pełni uchwycić. Problematyczny jest nierówny dostęp do informacji między stronami transakcji, utrudniający negocjacje. Sytuację dodatkowo komplikują niepewność rynkowa, zwłaszcza w sektorach technologicznych, oraz ryzyka prawne związane z błędną wyceną[2].

W styczniu 2022 r. Block, znany wcześniej jako Square, amerykański konglomerat technologii finansowej, zatwierdził emisję nowych akcji i zakończył przejęcie australijskiej firmy Afterpay za 29 mld USD. Ta transakcja, największa w sektorze BNPL (Kup teraz, zapłać później), miała na celu przekształcenie bazy klientów Afterpay w użytkowników aplikacji Cash App oraz wykorzystanie doświadczenia BNPL do ulepszenia swojego produktu [Shevlin, 2021].

Optymistyczne prognozy firmy Block dotyczące przyszłości firmy Afterpay okazały się jednak ryzykowne. S&P Global Market Intelligence wskazuje, że Block zapłacił znacznie zawyżoną cenę, wyceniając Afterpay na 29-krotność przychodów za następne 12 miesięcy – mnożniku znacznie wyższym niż w przypadku innych transakcji fin-

[2] Ryzyka prawne również przyczyniają się do niepewności wyceny. Wszelkie błędy lub przeoczenia w wycenie mogą prowadzić do sporów prawnych, wydłużając czas i zwiększając koszty transakcji M&A. Dodatkowo przewidywanie przyszłego wskaźnika wzrostu firmy jest z natury trudne, zwłaszcza w warunkach niepewności ekonomicznej, ze względu na oszacowanie długoterminowego wzrostu i założeń, na których oparte są prognozowane zwroty.

tech opartych na mnożniku przychodów historycznych[3] [Wang, Gull, 2021]. Zamiast zwiększać zyski dzięki Cash App, firma Afterpay miała trudności z utrzymaniem swojej wartości zakupu, co narażało firmę Block na wyższe ryzyko kredytowe, zwłaszcza w obliczu spowolnienia rynku fintech oraz globalnej recesji. Sytuacja ta podkreśla wyzwania związane z wycenami – ich dokładność zależy od założeń, a błędna wycena może podważyć wartość każdej transakcji M&A [Denning, 2016].

Tradycyjne metody, oparte na analizie finansowej, zdyskontowanych przepływach pieniężnych i porównaniach, mogą okazać się niewystarczające w przypadku szybko rozwijających się firm. Przykładem jest wycena firmy Afterpay, wykonana przez firmę Block i jej doradców, która opierała się na optymistycznych prognozach wartości niematerialnych i przepływów pieniężnych, co spotkało się z krytyką: „wycena Afterpay opierała się na założeniach o ciągłym wysokim wzroście przyszłych przepływów pieniężnych, a ten wysoki wzrost przyszłych przepływów pieniężnych wygląda mniej różowo" [Mullen, 2022]. W związku z tym coraz częściej do wycen wprowadza się nowoczesne technologie, takie jak sztuczna inteligencja i zaawansowana analiza danych. AI potrafi przetwarzać ogromne ilości nieustrukturyzowanych danych – raporty roczne, perspektywy branżowe, działania konkurentów, zachowania klientów itd., oferując bardziej kompleksowy i holistyczny obraz wartości firmy przejmowanej oraz wychwytując trendy i ryzyka, które mogą umknąć tradycyjnym metodom wyceny. Dzięki temu pracownicy i doradcy M&A mogą skupić się na elementach strategicznych, a nie na analizie danych, poprawiając tym samym efektywność procesu [Noghrehkar, 2024].

Szczególnie użyteczna jest analityka predykcyjna, wykorzystująca złożone modele algorytmiczne i historyczne dane do prognozowania przyszłych wyników. Dzięki temu firmy mogą precyzyjniej przewidywać przyszłe przepływy pieniężne i zmienne rynkowe. Block mógł skorzystać z tych narzędzi przy wycenie firmy Afterpay, aby uwzględnić więcej czynników wpływających na wartość: większą liczbę zmiennych finansowych i operacyjnych, a także sentyment konsumentów i popyt oraz trendy rynkowe. Modele wyceny oparte na AI są w stanie aktualizować dane w czasie rzeczywistym, uwzględniając wiadomości, recenzje, media społecznościowe, dzięki czemu stanowią niezwykle cenne narzędzie na dynamicznie zmieniających się rynkach. Taki model mógłby pomóc firmie Block w dokładniejszej wycenie Afterpaya, m.in. poprzez uwzględnienie aktualnych trendów rynkowych w sektorze BNPL[4].

---

3 "The deal values Afterpay at 29.29x next-12-months revenue rather than against trailing revenue or earnings, a multiple that is significantly higher than that of other fintech deals over the past few years. The elevated price is a sign that Square sees a huge market opportunity for BNPL going forward" [Wang, Gull, 2021].

4 Obecnie istnieje już kilku dostawców, oferujących różne modele wyceny oparte na AI. Jednak większość obecnych rozwiązań rynkowych jest skierowana na wyceny start-upów, co może nie być odpowiednie dla celów takich jak Afterpay.

Jednak wykorzystanie AI wiąże się z ryzykiem. Jakość wyników zależy od jakości danych wejściowych, a niekompletne lub stronnicze dane mogą prowadzić do błędnych wycen [Manyika, Silberg, Presten, 2019]. Istnieje problem „czarnej skrzynki", czyli interpretacji wyników: co dokładnie wpływa na wartość oraz czy wycena jest bezstronna. Jest to decydujące dla utrzymania zaufania i integralności procesu. AI wymaga również ciągłego monitorowania, aby zachować wiarygodność prognoz. Kolejne ryzyko odnosi się do ochrony danych, zwłaszcza w poufnych transakcjach M&A i przy zaangażowaniu zewnętrznych rozwiązań technologicznych.

Technologia AI, pod warunkiem odpowiedniego monitorowania oraz świadomości potencjalnych uprzedzeń w analizach, mogłaby wspomóc firmę Block w skuteczniejszym uwzględnianiu zmieniających się trendów rynkowych w modelu wyceny, a tym samym zmniejszyć ryzyko przepłacenia za przejmowaną spółkę. Wdrożenie jej wiąże się jednak z kosztami, które mogą stanowić barierę dla mniejszych firm. Nawet w przypadku firm, które stać na nowoczesne rozwiązania do wyceny, nie powinny one całkowicie zastępować ludzkiego osądu. Aby skutecznie korzystać z technologii, istotne jest łączenie wycen opartych na AI z tradycyjnymi modelami i analizą ekspertów. Balans między technologią a ekspercką wiedzą pozwala uzyskać wiarygodniejsze wyniki wyceny firm tradycyjnych oraz innowacyjnych, jak np. Afterpay, i podejmować lepsze decyzje w procesach M&A.

13.4. Zastosowanie technologii w integracji potransakcyjnej

Proces M&A jest pełen wyzwań, a każda faza niesie różne ryzyka, które mogą zmniejszyć wartość transakcji. Badania pokazują, że 70–90% transakcji kończy się niepowodzeniem, z czego znaczna część wynika z problemów w fazie integracji po fuzji [Kenny, 2020]. Pomimo lepszej analizy due diligence niemal połowa wszystkich działań integracji kończy się niepowodzeniem, zwiększając ryzyko dla przyszłych wyników [Gerds, Strottman, 2020]. Priorytetowe wyzwania obejmują utrzymanie tempa, skalowalność, różnice kulturowe, migrację danych, integrację technologii i systemów. W firmach opartych na danych szczególnie istotne są wyzwania, związane z ich niespójnością, jakością, zgodnością z przepisami oraz bezpieczeństwem. Migracja danych klientów jest szczególnie złożona, wymaga bowiem dostosowania różnych produktów i struktur oraz procedur identyfikacji klienta.

W styczniu 2022 r. brytyjska firma Farfetch, specjalizująca się w produkcji luksusowej odzieży, przejęła detalistę kosmetycznego Violet Grey za ponad 50 mln USD, wchodząc na rynek luksusowych kosmetyków. Dzisiejsi konsumenci oczekują holi-

stycznych doświadczeń, które odzwierciedlają ich podejście do *wellness* i indywidualności [Amed, Berg, 2023]. Przejęcie to miało strategicznie zwiększyć obecność Farfetch na rynku kosmetyków, choć integracja po fuzji przyniosła wyzwania, takie jak połączenie tożsamości marek oraz platform technologicznych. Chociaż Farfetch dążyło do długoterminowego wzrostu, dwa lata po przejęciu firma ogłosiła zamknięcie swojego działu kosmetycznego i wystawiła firmę Violet Grey na sprzedaż.

Integracja Farfetch i Violet Grey napotkała problemy związane z przyspieszonym harmonogramem, wyzwaniami operacyjnymi oraz przeszacowaniem siły marki Violet Grey w branży kosmetycznej. Szybkie tempo integracji (od przejęcia w styczniu 2022 r. do uruchomienia segmentu kosmetycznego w trzy miesiące) wzbudziło wątpliwości dotyczące staranności due diligence oraz dopasowania strategicznego. Problemy pojawiły się w logistyce, dostawach i zwrotach, a także w spełnianiu oczekiwań konsumentów co do szybkiej realizacji zamówień [Warn, 2023]. Ponadto Farfetch mogło przecenić wpływ swojego sukcesu w branży mody na działania na rynku kosmetycznym, ignorując fakt, że wielu konsumentów kosmetyków woli osobiście testować produkty przed zakupem [Strugatz, 2023].

W obliczu tych wyzwań technologie mogłyby pomóc w integracji operacyjnej, rozwoju marki w kosmetykach i zarządzaniu przyspieszoną integracją. Farfetch mógł wykorzystać narzędzia oparte na sztucznej inteligencji, takie jak IBM Watson, do analizy danych i identyfikacji problemów operacyjnych oraz trendów konsumenckich [IBM, 2024]. Analiza predykcyjna mogłaby poprawić zarządzanie zwrotami kosmetyków i kontrolą zapasów, a personalizacja oparta na AI mogłaby podnieść jakość doświadczeń zakupowych i zbudować zaufanie wśród klientów. Wdrożenie narzędzi RPA mogłoby usprawnić rutynowe zadania integracyjne, takie jak transfer danych, przyspieszając proces integracji bez uszczerbku dla jego precyzji i jakości oraz uwalniając zasoby ludzkie do bardziej skomplikowanych działań, jak np. dostosowanie strategii biznesowych łączących się firm.

Wdrożenie technologii wiąże się jednak z kosztami, które należy starannie rozważyć. Wysokie nakłady na AI i RPA wymagają dokładnego planowania finansowego, a same technologie mogą prowadzić do redukcji zatrudnienia oraz konieczności dodatkowego szkolenia personelu. Opór ze strony pracowników przyzwyczajonych do tradycyjnych metod może utrudniać wprowadzenie technologicznych rozwiązań. Dlatego integracja wymaga starannego zarządzania zmianą i utrzymania równowagi między wykorzystaniem technologii i czynnika ludzkiego, czego oczekują konsumenci firm, takich jak Farfetch.

Technologie usprawniają integrację danych, efektywnie łącząc systemy obu firm. Istnieją rozwiązania, takie jak platformy Mulesoft czy Boomi, które wykorzystują API

(interfejs programistyczny aplikacji), integrację w chmurze i RPA do sprawnego połączenia systemów, aplikacji i danych po fuzji. Mogą one ułatwić połączenie procesów firmy przejmowanej z procesami firmy przejmującej, a także zarządzanie i konsolidację danych, poprawiając ich jakość i dostępność w całym przedsiębiorstwie. Dodatkowo zapewniają zgodność danych z przepisami, takimi jak PSD2 i RODO. Wykorzystanie tych narzędzi zmniejsza ryzyko kar, zwiększa efektywność operacyjną, eliminując potrzebę ręcznych interwencji. Rozwiązania transformacji cyfrowej niezwykle ułatwiają firmom, takim jak SoFi czy Block, które regularnie przejmują inne przedsiębiorstwa, osiągnąć logiczną i efektywną infrastrukturę IT w sposób skalowalny.

Korzystanie z wymienionych rozwiązań wiąże się z wyzwaniami, takimi jak różnice kulturowe i organizacyjne, szczególnie gdy firmy pochodzą z różnych rynków, prowadząc do opóźnień w integracji. W trakcie procesu integracji mogą pojawić się konflikty między systemami danych i IT, wynikające z odmiennych sposobów podejmowania decyzji (centralizacja vs. decentralizacja, bezpośrednia vs. pośrednia komunikacja), zwiększające złożoność integracji. Dodatkowo wdrażanie zewnętrznych platform może zwiększać ryzyko związane z bezpieczeństwem i prywatnością danych.

Włączenie technologii do procesów integracji po fuzji jest fundamentalne, szczególnie w dynamicznym środowisku biznesowym. Dobrze zaplanowana integracja technologii, takich jak AI i RPA, może znacznie poprawić efektywność i bezpieczeństwo funkcjonowania połączonych organizacji. W przypadku dużych przejęć istotne jest staranne planowanie, testowanie narzędzi oraz monitorowanie postępów, aby zapewnić optymalne działanie połączonej infrastruktury. Jednak nawet mniejsze transakcje, jak przejęcie firmy Violet Grey przez firmę Farfetch, wymagają dokładnej oceny strategicznej, wykraczającej poza samo wdrożenie technologii. Trudności firmy Farfetch związane z wejściem na rynek kosmetyczny wskazują, że należy dokładnie przemyśleć, jak zintegrować firmę przejmowaną, aby zapewnić stabilność nowego segmentu biznesowego. Uniknięcie niepowodzeń i braku synergii, które mogłyby zagrozić powodzeniu przejęcia, ma znaczenie strategiczne. Technologia stanowi jedynie jeden z wielu elementów, mogących przyczynić się do udanej fuzji.

Podsumowanie

Co nie zadziałało w transakcjach SoFi-Technisys, HP-Autonomy, Block-Afterpay, Farfetch-Violet Grey i wielu innych fuzji i przejęć? Przyczyny niepowodzeń M&A są różne, ale wielu problemom można zaradzić, wykorzystując zaawansowane algo-

rytmy sztucznej inteligencji, analizę dużych zbiorów danych i hiperautomatyzację na poszczególnych etapach procesu M&A.

Analizowane przykłady transakcji M&A pokazują, że tradycyjne podejścia do fuzji i przejęć nie spełniają już wymagań nowoczesnych firm. Celem menadżerów i akcjonariuszy jest zwiększanie wartości firmy przez wzrost organiczny lub przejmowanie innych podmiotów. Podobnie jak nowoczesne technologie mogą zabezpieczać inwestycje kapitałowe w akcje spółek publicznych [Roszkowska, 2021], AI i hiperautomatyzacja mogą być implementowane na poszczególnych etapach procesu inwestowania w podmioty prywatne. Zastosowanie technologii poprawia efektywność podejmowania decyzji i w konsekwencji pomaga kreować wartość dla akcjonariuszy. Transakcje M&A są z natury unikatowe i w dużej mierze opierają się na ludzkim osądzie, który musi być obecny na każdym etapie transakcji. Aby jednak nadążyć za szybko ewoluującym środowiskiem biznesowym i zachować konkurencyjność, spółki przejmujące i doradcy muszą zacząć posiłkować się technologiami. Wykorzystanie AI i hiperautomatyzacji w M&A nie tylko zwiększa szanse powodzenia transakcji, ale także pokazuje, że transformacja cyfrowa może przynieść korzyści obszarom tradycyjnie zdominowanym przez ludzką ocenę i decyzje.

Bibliografia

Amed, I., Berg, A. (2023). *The State of Fashion Special*, https://www.businessoffashion.com/reports/beauty/the-state-of-fashion-beauty-industry-report-special-edition-2023-bof-mckinsey/ (dostęp: 28.08.2024).

Bakumenko, A., Elragal, A. (2022). Detecting Anomalies in Financial Data Using Machine Learning Algorithms, *Systems*, 10(5), s. 181–201.

Bobelian, M. (2012). *HP's $8.8 Billion Fiasco Exposes Flaws in M&A Process*, https://www.forbes.com/sites/michaelbobelian/2012/11/21/hps-8-8-billion-fiasco-exposes-flaws-in-ma-process/ (dostęp: 22.06.2024).

Clarke, S., Uhlaner, R., Wol, L. (2020). *A Blueprint for M&A Success*, https://www.mckinsey.com/capabilities/strategy-and-corporate-finance/our-insights/a-blueprint-for-m-and-a-success (dostęp: 8.08.2024).

Denning, S. (2016). *Valuation and the Case of False Precision*, https://www.forbes.com/sites/stevedenning/2016/08/19/valuation-and-the-case-of-false-precision/ (dostęp: 21.07.2024).

Deloitte (2021). *Due Diligence for Mergers and Acquisitions Through a Cybersecurity Lens*, https://www.deloitte.com/global/en/services/risk-advisory/blogs/due-diligence-for-mergers-and-acquisitions-through-a-cybersecurity-lens.html (dostęp: 4.10.2024).

Dilmegani, C. (2023). *RPA Due Diligence: Top 8 Use Cases in 2023*, https://research.aimultiple.com/rpa-due-diligence/ (dostęp: 23.06.2024).

Gerds, J., Strottman, F. (2020). *Post-Merger Integration: Hard Data, Hard Truths*, https://www2.deloitte.com/content/dam/insights/us/articles/post-merger-integration-hard-data-hard-truths/US_deloittereview_Post_Merger_Integration_Jan10.pdf (dostęp: 2.09.2024).

Hanelt, A., Firk, S., Hildebrandt, B., Kolbe, L.M. (2021). Digital M&A, Digital Innovation, and Firm Performance: An Empirical Investigation, *European Journal of Information Systems*, 30(1), s. 3–26.

IBM (2024). *Robotic Process Automation*, https://www.ibm.com/products/robotic-process-automation (dostęp: 27.08.2024).

Kenny, G. (2020). *Don't Make This Common M&A Mistake*, https://hbr.org/2020/03/dont-make-this-common-ma-mistake (dostęp: 2.09.2024).

Krohn, P. (2023). *AI: The Future of M&A*, https://middlemarketgrowth.org/dealmaker-ai-m-a-future/ (dostęp: 10.09.2024).

Manyika, J., Silberg, J., Presten, B. (2019). *What Do We Do About the Biases in AI?*, https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai (dostęp: 22.07.2024).

Moeller, S., Brady, C. (2014). *Intelligent M&A: Navigating the Mergers and Acquisitions Minefield*. Hoboken: John Wiley.

Mullen, C. (2022). *Will a Swoon in Valuations Affect Block's Afterpay?*, https://www.paymentsdive.com/news/will-blocks-afterpay-bet-pay-off-block-acquisition-buy-now-pay-later/628592/ (dostęp: 21.07.2024).

Müller, P.M., Päuser, P., Kurzrock, B.M. (2021). Fundamentals for Automating Due Diligence Processes in Property Transactions, *Journal of Property Investment and Finance*, 39(2), s. 97–124.

Noghrehkar, N. (2024). *Artificial Intelligence Tools for M&A Valuation*, https://imaa-institute.org/blog/ai-for-m-and-a- valuation-and-pricing/ (dostęp: 23.07.2024).

Patel, K. (2022). *10 Benefits of Mergers and Acquisitions You Should Know*, https://dealroom.net/blog/benefits-of-mergers-and-acquisitions (dostęp: 12.08.2024).

Roszkowska, P. (2021). Fintech in Financial Reporting and Audit for Fraud Prevention and Safeguarding Equity Investments, *Journal of Accounting and Organizational Change*, 17(2), s. 164–196.

Shevlin, R. (2021). *Why Square Acquired Afterpay for $29 Billion: Merchant and CashApp User Acquisition*, https://www.forbes.com/sites/ronshevlin/2021/08/02/why-square-acquired-afterpay-for-29-billion-merchant-and-cashapp-user-acquisition/ (dostęp: 21.07.2024).

SoFi Technologies (2022). *SoFi Completes Acquisition of Technisys*, https://investors.sofi.com/news/news-details/2022/SoFi-Completes-Acquisition-of-Technisys/default.aspx (dostęp: 24.07.2024).

SourceScrub (2024). *Best Sourcing Strategies*, https://www.sourcescrub.com/uk/post/best-sourcing-strategies (dostęp: 24.07.2024).

Strugatz, R. (2023). *Why It's So Hard for Fashion Retailers to Succeed in Beauty*, https://www.businessoffashion.com/articles/beauty/fashion-retailers-launch-beauty/ (dostęp: 28.08.2024).

Wang, Y., Gull, Z. (2021). *Square Pays up for Afterpay but Price Is Justified, Analysts Say*, https://www.spglobal.com/marketintelligence/en/news-insights/latest-news-headlines/square-pays-up-for-afterpay-but-price-is-justified-analysts-say-65827691 (dostęp: 21.07.2024).

Warn, G. (2023). *Why Is Farfetch Closing Beauty? 3 Ecommerce Experts Weigh*, https://britishbeautycouncil.com/why-is-farfetch-pulling-beauty/ (dostęp: 28.08.2024).

Ziuznys, A. (2022). *Deal Sourcing in 2024: Ultimate Guide*, https://coresignal.com/blog/deal-sourcing/ (dostęp: 10.08.2024).

14.0

GENERATYWNA SZTUCZNA INTELIGENCJA – NOWE WYZWANIA DLA RYNKU SZTUKI

Włodzimierz Szpringer
Szkoła Główna Handlowa w Warszawie

Streszczenie

Celem opracowania jest określenie wyzwań dla rynku sztuki, wynikających z wykorzystania generatywnej sztucznej inteligencji (GenAI). Potencjalne korzyści GenAI dla artystów są ogromne. Zagrożeniem może być konkurencja ze strony GenAI, gdyż może tworzyć sztukę, którą trudno odróżnić od tej stworzonej przez człowieka. Na tym tle dochodzi do sporów, czy utwory stworzone przy udziale GenAI nie naruszają praw innych osób, gdy GenAI kopiuje dzieła (lub style) innych twórców lub przekształca teksty na obraz lub dźwięk. Konflikty wynikają również w związku z eksploracją tekstów i danych (Text and Data Mining, TDM), niezbędnych do szkolenia GenAI.

Słowa kluczowe: generatywna sztuczna inteligencja, sektory kreatywne, rynek sztuki, transformacja cyfrowa, zarządzanie technologią

Wprowadzenie

Istnieją sprzeczne poglądy na temat roli sztucznej inteligencji w branżach kreatywnych. Dla niektórych jest to szansa, dla innych zagrożenie. Z jednej strony istnieje ryzyko, że niektóre osoby stracą pracę w wyniku automatyzacji opartej na AI. Z drugiej

zaś okazuje się, że AI jest bardzo przydatna, ponieważ ułatwia stworzenie innowacyjnego utworu czy to w muzyce, malarstwie, a nawet poezji. Należy zatem spojrzeć na pozytywne aspekty AI i zobaczyć, jak inspirująca może być dla ludzi. Branża ta ma potencjał rozwoju i jest to najlepszy czas, aby skorzystać ze sztucznej inteligencji [Tesseract, 2023]. Między tradycyjnymi artystami a tymi, którzy tworzą treści AI, nadal istnieje pewne napięcie. Niektórzy obawiają się, że rozwój technologii sztucznej inteligencji zastąpi ludzką kreatywność i doprowadzi do ujednolicenia sztuki i kultury. Inni mówią, że jest odwrotnie. „Demokratyzując" proces twórczy i czyniąc go bardziej dostępnym, technologia AI umożliwia usłyszenie większej liczby głosów i udostępnienie większej liczby perspektyw [Guler, 2023; Hayden, 2023].

Na ChatGPT nie należy patrzeć przez pryzmat binarności, a raczej jak na ludzką kreatywność, wspomaganą sztuczną inteligencją, tym bardziej że współpraca między technologią uczenia maszynowego a ludźmi może wkrótce stać się normą. Należy tak sformułować tę narrację, aby unikać wypowiedzi: „sztuczna inteligencja albo człowiek", a mówić: „sztuczna inteligencja i człowiek". Współpraca człowieka i AI może bowiem często skutkować pracą, która jest lepsza od tych niewzmocnionych przez AI, przy jednoczesnym zachowaniu istoty ludzkiej kreatywności. Człowiek „plus" sztuczna inteligencja to potężniejszy silnik kreatywny ze względu na możliwą do uzyskania wydajność, uzyskaną dzięki wiedzy, jaką możemy posiadać np. o muzyce, o rozwijaniu i rozumieniu dźwięków, czy mieszaniu się kultur i tworzeniu nowych dźwięków. Tak właśnie dzieje się każdego dnia w branży muzycznej [Sperling, 2023].

14.1. Problem definiowania kreatywności – czy AI może być twórcą?

Nowy model kreatywności z wykorzystaniem AI znacznie obciąża dwie najbardziej podstawowe doktryny prawne prawa autorskiego: dychotomię idei i ekspresji tych idei oraz test istotnego podobieństwa w przypadku jego naruszenia. Coraz większa kreatywność będzie polegać na zadawaniu właściwych pytań, a nie na tworzeniu odpowiedzi. Zadawanie pytań może być kreatywne, a AI wykonuje większość pracy, za którą dotychczas nagradzano prawa autorskie, i praca ta nie jest chroniona. Następuje odwrócenie tego, co obecnie ceni prawo autorskie. Skoro zadawanie pytań byłoby podstawą zdolności autorskiej, podobieństwo wyrażeń w odpowiedziach nie będzie już bardzo przydatne w udowadnianiu faktu skopiowania pytań. Oznacza to, że być może trzeba będzie odrzucić test na naruszenie lub przynajmniej zastosować go w zasadniczo odmienny sposób [Lemley, 2023].

Podobnie jak ludzie przywiązują wagę do interakcji międzyludzkich w branżach kreatywnych, w przestrzeni innowacji ceni się produkty stworzone przez człowieka. Dzięki świadomości, że dany produkt stworzyła prawdziwa osoba, prawdopodobnie uzyskałby on wyższą cenę na rynku. GenAI nadal może mieć znaczenie we wprowadzaniu na rynek nowych produktów, np. poprzez podsumowanie badań rynkowych i nowych spostrzeżeń konsumentów. Ludzie jednak nadal mają do odegrania ważną rolę na rynku kreatywnym. Inspiracją do kreatywności jest życie, rzeczy, które nam się przytrafiają. Ale sztuczna inteligencja „nie ma życia", więc nie dzieją się w niej przypadkowe rzeczy, które inspirują kreatywność. Chociaż AI może pomóc w generowaniu nowatorskiego materiału, jej wartość estetyczna może nadal być ograniczona, ponieważ opiera się na istniejących danych i wzorcach.

Dzieło może być wyjątkowe w tym sensie, że nigdy wcześniej nie powstało, ale może brakować mu głębi i kreatywności, jakie artyści wnoszą do swojej pracy. Sztuczna inteligencja pomogła już w pisaniu piosenek, naśladowała styl wielkich malarzy i podejmowała decyzje twórcze podczas kręcenia filmów. Eksperci zastanawiają się jednak, jak daleko AI może lub powinna sięgać w procesie twórczym. Opinie na temat, czy AI ma potencjał, aby stać się prawdziwym partnerem twórczym, a nawet twórcą solowych dzieł sztuki, są różne. Chociaż debata ta prawdopodobnie będzie trwać przez jakiś czas, jasne jest, że w miarę, jak treści cyfrowe i platformy ich dostarczania będą w dalszym ciągu infiltrować wszystkie formy mediów i ekspresji, rola AI niewątpliwie wzrośnie. Głębokie uczenie się nie jest odpowiedzią na kreatywność [IBM, 2023].

Należy jednak zdefiniować, co oznacza kreatywność. Wiemy, że niektóre atrybuty mają związek ze znalezieniem czegoś nowego, nieoczekiwanego, a jednocześnie przydatnego. Inspiracja to jedna z ról, jaką sztuczna inteligencja może odgrywać w procesie twórczym, ale może również pomóc w bardziej przyziemnych zadaniach, szczególnie w domenie cyfrowej, gdzie większość pracy nie jest efektowna. Ostatnio w świecie GenAI dokonano wielu postępów. Jeśli chodzi o kreatywność, nie można zaprzeczyć, że AI będzie miała trudności z pełnym zastąpieniem ról twórczych, zwłaszcza że sztuka i projektowanie są bardzo subiektywne, ale istniejące już narzędzia posiadają zdolności do tworzenia złożonych dzieł sztuki, które mogą wywoływać ludzkie emocje. W obliczu zmian technologicznych kreatywność jest często postrzegana jako cecha typowo ludzka, mniej podatna na siły zakłóceń technologicznych i mająca kluczowe znaczenie dla przyszłości. Generatywne aplikacje AI, takie jak ChatGPT czy Midjourney, stawiają jednak na nowo pytania o definicję kreatywności.

Już dziś generatywne aplikacje AI grożą znaczącymi zmianami w pracy twórczej, zarówno niezależnej, jak i płatnej. Te nowe modele GenAI uczą się na podstawie ogromnych zbiorów danych i opinii użytkowników i mogą tworzyć nowe treści

w postaci tekstu, obrazów oraz dźwięku lub ich kombinacji. W związku z tym wydaje się, że obecnie zawody skupiające się na dostarczaniu treści, takich jak pisanie, tworzenie obrazów, kodowanie, a także tych, które zazwyczaj wymagają dużej ilości wiedzy i informacji, będą prawdopodobnie szczególnie inspirowane przez GenAI. Dzięki nowym kanałom cyfrowym niezależni autorzy, podcasterzy, artyści i muzycy mogą bezpośrednio kontaktować się z odbiorcami w celu uzyskania własnych dochodów. Platformy internetowe, takie jak Substack, Flipboard i Steemit, umożliwiają jednostkom nie tylko tworzenie treści, ale również stanie się niezależnymi producentami i menadżerami marki swojej pracy. Chociaż nowe technologie zakłócały wiele rodzajów pracy, platformy te oferują ludziom nowe sposoby zarabiania na życie dzięki ludzkiej kreatywności [De Cremer, Morini Bianzino, Falk, 2023].

14.2. Przykłady „twórczości" w kontekście GenAI

Przyjrzyjmy się przykładom „twórczości" w kontekście GenAI. Jednym z nich jest Midjourney, niezależne laboratorium badawcze, autorski program sztucznej inteligencji, który tworzy obrazy z opisów tekstowych, podobnie jak DALL-E OpenAI. Kolejne dwie firmy, IBM Watson i Sony Flow Machine, stanowią przykłady wykorzystania mocy sztucznej inteligencji do tworzenia muzyki. Popularny producent muzyczny Alex Kid połączył siły z programem Watson z IBM, aby wspólnie stworzyć piosenkę. Została ona w całości napisana i wyprodukowana przy użyciu AI. Sztuczną inteligencję w świecie muzyki zastosowano również „na żywo" podczas wirtualnych koncertów, organizowanych przez muzyków dla tysięcy ludzi. Przykładem jest zespół ABBA i ich całkowicie wirtualna trasa koncertowa Voyage. Zespół nawiązał współpracę z Industrial Light and Magic, firmą zajmującą się efektami wizualnymi, która wykorzystała zaawansowane techniki przechwytywania ruchu do stworzenia wirtualnych kopii zespołu, które zachowują się dokładnie tak samo jak członkowie zespołu, naśladując ich ruchy oczu i ruchy taneczne [Beyond, 2023].

DALL-E 2 – to system AI, który tworzy realistyczne obrazy artystyczne na podstawie poleceń w języku naturalnym. Interpretuje język, a następnie daje twórcze rezultaty. Może także rozszerzyć już istniejące dzieła sztuki lub zmienić ich przeznaczenie. AdCreative – to platforma AI, która umożliwia tworzenie i przesyłanie wszystkich treści w mediach społecznościowych zawierających dane oparte na sztucznej inteligencji. Czy AI nie jest w stanie napisać całych powieści? Początkowo maszyny mogły pisać jedynie krótkie treści w stylu „dziennikarskim", ale obecnie pojawiło się kilka przykładów wykorzystania AI do pisania całych powieści. Wyniki były nieco wątpli-

we i chociaż tekst czyta się jak powieść, nie jest to to samo, co powieść napisana przez człowieka. Sztuczna inteligencja poczyniła znaczne postępy w automatyzacji niektórych aspektów projektowania graficznego, takich jak generowanie zmian w projekcie, analizowanie danych i pomoc w powtarzalnych zadaniach. Godnym uwagi przykładem jest sztuczna inteligencja Canvas Design, która wykorzystuje algorytmy uczenia maszynowego, aby pomóc użytkownikom projektować atrakcyjną wizualnie grafikę. Sugeruje elementy projektu, kombinacje kolorów i opcje układu w oparciu na preferencjach użytkownika i trendach projektowych. Chociaż Design AI usprawnia proces projektowania, nadal wymaga wkładu człowieka i podejmowania decyzji, aby nadać ostatecznie projektowi osobisty charakter i kreatywną wizję [Logan, 2023].

Aktorstwo to zawód, który w dużej mierze opiera się na ludzkich emocjach, ekspresji i interpretacji. Chociaż AI nie jest w stanie odtworzyć głębi ludzkich emocji, znalazła pewne zastosowania w branży aktorskiej. Narzędzia oparte na sztucznej inteligencji, takie jak chatGPT, zostały wykorzystane do generowania skryptów i dialogów, a nawet pomagają w rozwoju postaci. Powstaje pytanie, w jaki sposób AI została wykorzystana do tworzenia „wirtualnych aktorów" w grach wideo i filmach. Ci wirtualni aktorzy mogą dostosowywać swoje występy na podstawie opinii widzów w czasie rzeczywistym, tworząc bardziej wciągające doświadczenia. Należy jednak zauważyć, że występom generowanym przez AI brakuje spontaniczności, niuansów i umiejętności improwizacji, które aktorzy wnoszą do swojego rzemiosła. Sztuczna inteligencja okazała się obiecująca w branży aktorskiej, pomagając w generowaniu scenariusza i rozwoju postaci. Może analizować ogromne ilości danych, takich jak scenariusze, filmy i spektakle, aby identyfikować wzorce i generować nowe pomysły. Algorytmy AI mogą dostarczyć cennych informacji na temat rozwoju postaci, pomagając aktorom lepiej zrozumieć swoje role. Jednak ostateczna interpretacja i przedstawienie postaci wymagają kreatywności, emocji i intuicji, które są cechą ludzi.

Sztuczna inteligencja poczyniła także postępy w komponowaniu muzyki. Analizując istniejące kompozycje i wzorce muzyczne, algorytmy AI mogą generować oryginalne melodie i stanowić cenne narzędzie dla kompozytorów, dostarczające im nowych pomysłów i inspiracji. AI nie może jednak odtworzyć głębi emocjonalnej, umiejętności opowiadania historii i osobistych doświadczeń, jakie kompozytorzy wnoszą do swojej twórczości. Muzyka jest głęboko ludzką formą sztuki, a emocje, jakie wywołuje, wynikają z osobistego i niepowtarzalnego punktu widzenia jej twórców.

W fotografii sztuczną inteligencję wykorzystuje się do poprawy jakości obrazu, automatyzacji zadań przetwarzania końcowego, a nawet generowania realistycznych obrazów. Algorytmy AI mogą analizować ogromne zbiory danych zdjęć, aby uczyć się wzorców i generować estetyczne obrazy. Chociaż może to być przydatne narzędzie

dla fotografów, umiejętność uchwycenia emocji, opowiadania historii i tworzenia związku z obiektem wymaga ludzkiego oka, intuicji i empatii.

W miarę jak sztuczna inteligencja staje się coraz zdolniejsza do naśladowania głosów i stylów znanych artystów, w grę wchodzą względy autorskie i etyczne. Przypadek utworu wygenerowanego przez AI, w którym pojawia się głos Drake'a, rodzi pytania o naruszenie praw autorskich i własność dzieł kreatywnych. Chociaż sztuczna inteligencja mogłaby zostać wykorzystana jako narzędzie do stworzenia czegoś nowego, to w artyście należy cenić kreatywność i oryginalność. Podobnie w branży filmowej wykorzystanie AI do odtwarzania podobizny aktorów wywołało dyskusje na temat praw wykonawców oraz roli reżyserów i producentów jako artystów kreatywnych.

Aby zapewnić poszanowanie praw i wkładu artystów, należy wziąć pod uwagę prawne i etyczne konsekwencje stosowania AI w procesach twórczych. Sztuczna inteligencja będzie nadal odgrywać znaczącą rolę w tworzeniu sztuki. Ważne jest jednak, aby rozpoznać ograniczenia AI i wartość, jaką artyści-ludzie wnoszą do procesu twórczego. Sztuczna inteligencja może być potężnym narzędziem, które wzmacnia i poszerza możliwości artystyczne, ale nie może zastąpić głębi, kreatywności i wyjątkowej perspektywy, jaką oferują artyści. W miarę postępu technologii AI zaistniała też potrzeba ciągłych dyskusji i współpracy między artystami, technologami i decydentami, aby zająć się prawnymi, etycznymi i społecznymi konsekwencjami sztucznej inteligencji w twórczości artystycznej.

Wielcy artyści tak jak The Beatles podjęli już eksperymenty ze sztuczną inteligencją. Ostatnia „nowa" piosenka zespołu *Now And Then* została stworzona przy użyciu AI, aby ożywić niewydane nagranie demo zmarłego Johna Lennona. Pozostali dwaj członkowie zespołu, Paul McCartney i Ringo Starr, w listopadzie 2023 r. wydali singiel *Now And Then*. To ich ostatnia piosenka, w której wystąpili wszyscy czterej Beatlesi, oraz ich pierwsza i jedyna piosenka wydana w XXI w. Korzystając ze sztucznej inteligencji, muzycy i autorzy tekstów mają możliwość generowania treści w ciągu kilku sekund, syntezy wokali o podobnym brzmieniu, oddzielania elementów w tym samym utworze i wiele więcej. W przypadku *Teraz i wtedy* zespołu The Beatles AI miała przede wszystkim charakter regeneracyjny, a nie generatywny. Oprogramowanie stworzone przez zespół Petera Jacksona podczas kręcenia filmu dokumentalnego *The Beatles: Get Back* wykorzystywało AI do izolowania głosu Johna Lennona z nagrania, tak przeplatającego się z hałasem i fortepianem, że wcześniej nie nadawało się do użytku. Sztuczna inteligencja stworzyła piosenkę *Dady's Car* w stylu The Beatles z tekstami napisanymi przez ludzi. AI przeanalizowała bazę danych piosenek Beatlesów, aby stworzyć podobną kompozycję. Piosenka nie do końca dorównuje Beatlesom, ale sztuczna inteligen-

cja tworząca piosenki na podstawie istniejących utworów starszych artystów może stać się bardziej powszechna [Ohio, 2023].

Inny problem, z którym boryka się już wiele osób w branżach kreatywnych, to bezprecedensowa ilość treści tworzonych przez ludzi trafiających do głównego nurtu. Codziennie na głównych platformach streamingowych, takich jak Spotify i Apple Music, publikowanych jest od 100 tys. do 150 tys. utworów. Sztuczna inteligencja jedynie przyspieszy liczbę utworów wprowadzanych na rynek. W nadchodzących latach staniemy w obliczu tsunami utworów generowanych przez AI, od muzyków-amatorów po czołowych artystów. Wiele platform do przesyłania strumieniowego wykorzystuje algorytmy oparte na AI do dostarczania muzyki użytkownikom. Dla słuchaczy odkrywanie nowych treści jest przytłaczające, a dla nowych artystów prawie niemożliwe staje się wyróżnienie się z tłumu. Autorzy, muzycy, aktorzy, filmowcy i inne osoby pracujące w branżach kreatywnych zdali sobie sprawę z zagrożenia, jakie dla ich pracy stanowi GenAI. W USA grupa 17 wybitnych autorów pozwała firmę technologiczną OpenAI, właściciela usługi ChatGPT, za „systematyczną kradzież na masową skalę". W pozwie zbiorowym zarzucają firmie nielegalne wykorzystywanie dzieł chronionych prawem autorskim oraz to, że jej chatbot może tworzyć nowe dzieła w stylu swoim i innych autorów bez ich zgody i bez udziału w zysku. Rzecznik OpenAI powiedział, że firma szanuje „prawa pisarzy i autorów i wierzy, że powinni oni korzystać z technologii sztucznej inteligencji" [Hodge, 2023].

Wyzwaniem jest znalezienie właściwej równowagi między zachęcaniem do innowacji a ochroną praw własności intelektualnej w erze sztucznej inteligencji. Zbyt duża ochrona może stłumić innowacje, natomiast zbyt mała może zniechęcić osoby fizyczne i przedsiębiorstwa do inwestowania w rozwój AI ze względu na brak wystarczającego zwrotu z inwestycji. Jest jednak wątpliwe, czy DABUS[1] można uznać za autonomicznego wynalazcę [Matulionyte, 2022].

Zatem w przypadkach, gdy ktoś jako wynalazcę podaje AI, proponuje się ustanowienie wysokiego standardu dowodu w postaci przedstawienia rzeczywistej wydajności maszyny zgłoszonej jako wynalazek oraz zilustrowania, w jaki sposób maszyna działa, aby osiągnąć taki wynik. Sztuczna inteligencja podnosi również kwestię potencjalnego naruszenia praw własności intelektualnej. Na przykład systemy sztucznej inteligencji są często szkolone przy użyciu dużych zbiorów danych, które mogą zawierać

[1] DABUS – to system sztucznej inteligencji stworzony do „generowania wynalazków" przez amerykańskiego wynalazcę Stephena Thalera. Sprawa dotyczyła dwóch zgłoszeń patentowych, dokonanych w Europejskim Urzędzie Patentowym przez twórcę DABUS-a w 2018 r. Izba Odwoławcza Europejskiego Urzędu Patentowego potwierdziła, że patent może uzyskać tylko osoba fizyczna. Podobnie orzekają sądy w innych krajach, z nielicznymi wyjątkami (RPA) [Czajkowska, Walasek, 2022].

materiały chronione prawem autorskim. Jeżeli system AI generuje dane wyjściowe na ich podstawie, czy może to stanowić naruszenie praw autorskich? Odpowiedź zależy od efektu – np. stopnia podobieństwa do utworów bazowych, stopnia transformacji, kopiowania stylu itp. [Ghanghash, 2023].

14.3. Przykłady konfliktów na tle GenAI

Na tle GenAI pojawiły się konflikty, artyści wyrażali wątpliwości co do wykorzystania ich oryginalnych dzieł. Media i przemysł kreatywny były sondowane przez rząd w tej kwestii. Wśród twórców rosły obawy, że produkty generowane przez sztuczną inteligencję naruszają ich prawa autorskie. Dyrektorzy grup medialnych skarżyli się, że firmy technologiczne wykorzystują ich treści do szkolenia modeli LLM bez umowy licencyjnej. Inne podmioty działające w sektorach muzycznym, wydawniczym i nadawczym mają podobne wątpliwości co do korzystania z ich oryginalnych dzieł – np. podczas tworzenia „głęboko fałszywych" produkcji, generowanych przez sztuczną inteligencję z zastosowaniem głosu lub stylu artysty do tworzenia nowych dzieł. Pomiędzy firmami technologicznymi i medialnymi toczą się wstępne dyskusje na temat licencjonowania produktów generowanych przez sztuczną inteligencję w oparciu na materiałach chronionych prawem autorskim. Nowa instytucja eksploracji tekstów i danych (TDM), wprawdzie dość wąsko zakreślona, obarczona jest niepewnością prawną, która wywołuje konflikty na tle praw autorskich [FT, 2023].

W sprawie Andersen przeciwko Stability AI *et al.*, złożonej pod koniec 2022 r., trzech artystów utworzyło grupę, aby pozwać wiele platform GenAI ze względu na to, że ich oryginalne dzieła wykorzystywane są bez licencji na szkolenie sztucznej inteligencji w zakresie ich stylów, umożliwiając użytkownikom generowanie dzieł, które mogą w niewystarczającym stopniu przekształcać ich istniejące chronione dzieła, które w rezultacie byłyby nieautoryzowanymi dziełami pochodnymi. Jeżeli sąd uzna, że dzieła sztucznej inteligencji są nieautoryzowane i pochodne, może nałożyć poważne kary za naruszenie praw autorskich. Podobne sprawy złożone w 2023 r. prowadzą do twierdzeń, że firmy szkoliły narzędzia sztucznej inteligencji przy użyciu dużych zbiorów danych zawierających tysiące – a nawet wiele milionów – nielicencjonowanych dzieł. Getty Images, firma zajmująca się licencjonowaniem obrazów, złożyła pozew przeciwko twórcom Stable Diffusion, zarzucając im niewłaściwe wykorzystanie jej zdjęć, z naruszeniem zarówno praw autorskich, jak i praw do znaków towarowych przysługujących jej w zakresie kolekcji zdjęć ze znakiem wodnym. W każdym z tych przypadków system prawny proszony jest o wyjaśnienie granic, czym jest

„utwór pochodny" w rozumieniu przepisów dotyczących własności intelektualnej – przy czym w zależności od jurysdykcji różne sądy mogą przedstawiać różne interpretacje. Oczekuje się, że wynik tych spraw będzie zależał od interpretacji doktryny dozwolonego użytku, która pozwala na wykorzystywanie utworu chronionego prawem autorskim bez zgody właściciela „w celach takich jak krytyka (w tym satyra), komentowanie, relacjonowanie wiadomości, nauczanie (w tym wielokrotne kopie do użytku w klasie), lub badania" oraz do przekształcającego wykorzystania materiału chronionego prawem autorskim w sposób, do którego nie był przeznaczony [Appel, Neelbauer, Schweidel, 2023].

W pozwie złożonym przez „The New York Times" (NYT) zarzuca się, że Microsoft i OpenAI wykorzystały artykuły publicznie dostępne w witrynie czasopisma do stworzenia produktów sztucznej inteligencji, które konkurują z możliwością świadczenia przez gazetę internetowego serwisu informacyjnego i im zagrażają. Narzędzia generujące sztuczną inteligencję stosowane przez pozwanych opierają się na dużych modelach językowych, które zbudowano poprzez kopiowanie i wykorzystywanie milionów chronionych prawem autorskim artykułów prasowych, szczegółowych badań, opinii, recenzji, poradników i nie tylko. Gazeta NYT stwierdziła, że chociaż Microsoft i OpenAI zajmowały się kopiowaniem na szeroką skalę z wielu źródeł, to przy tworzeniu swoich modeli LLM położyły szczególny nacisk na treści pochodzące właśnie z NYT. Za pośrednictwem usługi Bing Chat firmy Microsoft (niedawno przemianowanej na Copilot) i ChatGPT firmy OpenAI oskarżeni próbują wykorzystać ogromną inwestycję w dziennikarstwo, stosując je do tworzenia produktów zastępczych bez pozwoleń i wynagrodzeń [Saran, 2024; Ajao, 2024].

W Wielkiej Brytanii firmie StabilityAI nie udało się odrzucić pewnych twierdzeń, że naruszyła ona prawa własności intelektualnej Getty Images, zanim sprawa trafiła do brytyjskiego sądu. Omawiając oba pozwy oraz sposób szkolenia modeli LLM, można skonstatować, że w badanych przypadkach istnieje przynajmniej element czytania, tworzenia kopii, a następnie uruchamiania robotów indeksujących lub sztucznej inteligencji w celu uczenia się. Wykonywanie kopii po drodze jest częścią procesu szkoleniowego. Czynność tworzenia kopii treści podlega jednak ograniczeniom na mocy praw autorskich. Jeśli dana sprawa nie zostanie objęta jednym z kilku wyjątków dotyczących praw autorskich, będzie to naruszenie, a trzeba dodać, że nie jest łatwo uzyskać tego rodzaju komercyjne ćwiczenia szkoleniowe dotyczące wszelkich wyjątków. Choć argumenty prawne mogą być inne w USA, Europie czy Wielkiej Brytanii, można założyć, że spośród różnych firm oferujących usługi GenAI przynajmniej część tej działalności prawdopodobnie narusza prawa własności intelektualnej. Getty Images zarabia na licencjonowaniu praw do korzystania z obrazów w swojej

ogromnej bibliotece obrazów. Obrazy te są często włączane do broszur firmowych i slajdów. Jeżeli dostawca usług GenAI dotyczących tworzenia obrazów może tworzyć treści podobne do obrazów w firmie zajmującej się biblioteką obrazów, podważa to model biznesowy tej ostatniej [Klovig, Skelton, 2024].

W 2023 r. wielu wydawców muzycznych, w tym Concord, Universal Music Group i ABKCO, wszczęło postępowanie prawne przeciwko wspieranej przez Amazon i Google firmie Anthropic, oferującej usługi zbliżone do chatGPT, żądając potencjalnie wielomilionowych odszkodowań za rzekomo systematyczne i powszechne naruszanie ich utworów chronionych prawem autorskim. W pozwie złożonym przed sądem rejonowym w Tennessee zarzucono, że firma Anthropic, budując i obsługując swoje modele sztucznej inteligencji, bezprawnie kopiuje i rozpowszechnia ogromne ilości dzieł chronionych prawem autorskim – w tym teksty niezliczonych kompozycji muzycznych, będących własnością wydawców lub przez nich kontrolowanych. Firma Anthropic, zajmująca się sztuczną inteligencją, ostro zareagowała na pozew dotyczący praw autorskich złożony przez wydawców muzycznych, twierdząc, że treści wykorzystane w jej modelach podlegają dozwolonemu użytkowi (*fair use*) i że jakikolwiek system licencjonowania byłby zbyt skomplikowany i kosztowny. Stwierdziła ponadto, że wykorzystywanie treści chronionych prawem autorskim w danych szkoleniowych dotyczących modelu dużego języka (LLM) mieści się w ramach dozwolonego użytku oraz że dzisiejsze narzędzia sztucznej inteligencji ogólnego przeznaczenia po prostu nie mogłyby istnieć, gdyby firmy AI musiały płacić za licencje. Zgodnie z prawem USA „dozwolony użytek" pozwala na ograniczone wykorzystanie materiałów chronionych prawem autorskim bez zgody autorów, do celów takich jak krytyka, reportaże prasowe, nauczanie i badania.

Chociaż technologia sztucznej inteligencji może być złożona i nowatorska, kwestie prawne związane z wykorzystaniem materiałów chronionych prawem autorskim są w gruncie rzeczy proste. Pozwany nie może reprodukować, rozpowszechniać ani wyświetlać cudzych dzieł chronionych prawem autorskim w celu budowania własnego biznesu, chyba że uzyska zgodę posiadacza praw. Zasada ta nie zanika po prostu dlatego, że przedsiębiorstwo „wzbogaca" swoje naruszenie słowem „AI". W zgłoszeniu stwierdzono ponadto, że niezapewnienie przez Anthropic zezwoleń na prawa autorskie pozbawia wydawców i ich autorów piosenek kontroli nad ich dziełami chronionymi prawem autorskim oraz ciężko wypracowanymi korzyściami z ich twórczych wysiłków. Aby załagodzić problem, wydawcy muzyczni wzywają sąd do nałożenia na Anthropic odszkodowania w postaci obowiązku zapewnienia rozliczania swoich danych, treści i metod szkoleniowych oraz zniszczenia wszystkich kopii naruszających prawo, będących w posiadaniu firmy.

Firma Anthropic stwierdziła jednak, że szkolenie jej modelu sztucznej inteligencji kwalifikuje się jako zasadniczo zgodne z prawem wykorzystanie materiałów, przywołując argumenty, że w zakresie, w jakim dzieła chronione prawem autorskim są stosowane w ramach danych szkoleniowych, służy to analizie (statystycznych powiązań między słowami i pojęciami), niezwiązanej z jakimkolwiek wyrazistym celem dzieła. Wykorzystywanie utworów do szkolenia tego systemu AI jest sprawiedliwe, ponieważ nie uniemożliwia sprzedaży oryginalnych dzieł, a nawet jeśli ma charakter komercyjny, nadal jest on wystarczająco transformacyjny wobec utworów chronionych prawnie. Jeśli chodzi o potencjał systemu licencyjnego, Anthropic argumentowała, że ciągłe wymaganie licencji byłoby niewłaściwe, ponieważ zablokowałoby dostęp do zdecydowanej większości utworów i przyniosłoby korzyści tylko podmiotom posiadającym największe zasoby, które są w stanie płacić za przestrzeganie tak interpretowanych przepisów. Wymaganie licencji na niewyrażone w sposób tradycyjny wykorzystanie dzieł chronionych prawem autorskim w celu szkolenia LLM oznacza utrudnianie wykorzystywania pomysłów, faktów i innych materiałów nieobjętych prawami autorskimi. Nawet przy założeniu, że pewne aspekty zbioru danych mogą przypisywać danemu wynikowi większą wagę niż inne, model jest czymś więcej niż tylko sumą jego części. W związku z tym trudno będzie ustalić stawkę tantiem, która byłaby istotna dla indywidualnych twórców, nie czyniąc przy tym przede wszystkim nieekonomicznym opracowywania generatywnych modeli sztucznej inteligencji.

Podsumowanie

Technologia cyfrowa znacząco zmienia równowagę w relacjach społecznych i gospodarczych, oferując nowe możliwości dla innowacyjnych modeli biznesowych. Rodzi to nowe pytania, które dotyczą wielu aspektów prawa, w tym np. klasyfikacji prawnej danych, eksploracji tekstu i danych, handlu danymi, powiązanych z danymi kwestii ochrony własności intelektualnej i dostarczania treści cyfrowych oraz wyzwań związanych z internetem rzeczy, sztuczną inteligencją i regulacją algorytmiczną. GenAI rodzi zarówno szanse, jak i zagrożenia dla rynku sztuki. Brak pewności prawnej stanowi jednak duże zagrożenie. Powstają pytania: jakie są granice transformacyjnej „przeróbki" dzieła, czy dzieło powstałe przy wykorzystaniu GenAI to utwór pochodny czy całkiem nowy? Komu należy przypisać autorstwo dzieła powstałego przy użyciu GenAI? W kwestii eksploracji tekstu i danych można czerpać inspiracje z początków streamingu, gdy serwisy takie jak Napster oferowały bezpłatne pliki do pobrania. Nikt tak naprawdę nie wiedział, co jest zgodne z prawem, a co nie. Wraz z wprowadzeniem

Spotify pojawił się legalny streaming – model licencjonowany. Wiedziano, że muzyka w Spotify jest bezpieczna z perspektywy zagrożeń wirusami komputerowymi, lepsze są także relacje z klientami. Świat sztucznej inteligencji może przejść przez to samo. Rozwój orzecznictwa sugeruje, że koncepcja komunikacji ewoluuje zgodnie ze zmieniającymi się sposobami udostępniania plików w epoce cyfrowej. W miarę ciągłego rozwoju technologii i sposobów udostępniania treści, umożliwiających łączenie dużych ilości mediów lub treści za pomocą procesu częściowo lub w pełni algorytmicznego, pośrednie etapy w łańcuchu komunikacji stają się coraz bardziej zautomatyzowane, co skutkuje redukcją poziomu aktywnego zaangażowania (czy interwencji wymaganej ze strony gospodarza lub użytkownika). „Test komunikacji", który tradycyjnie opierał się na bezpośredniej czynności udostępnienia treści, został w ostatnich latach znacznie złagodzony. Następuje stopniowe rozluźnianie standardów stosowanych do ustalania, czy prawo autorskie zostało naruszone. Tradycyjny standard „transmisji" był krytykowany jako zbyt rygorystyczny. Oznacza to dalsze obniżanie testu wykorzystywanego w celu ustalenia, czy akt komunikacji miał miejsce, co przybiera formę standardu „ułatwiającego", który zastępuje lub uzupełnia standard „interwencyjny". Rozszerzenia definicji komunikacji muszą być zgodne z charakterem i celem protokołów wymiany plików stosowanych w nowszych wersjach platform cyfrowych i powstających rynków handlu wirtualnego. Jesteśmy obecnie w początkowej fazie sporów sądowych, rozmów ugodowych i licencyjnych. Konieczne jest złagodzenie tej niepewności dzięki konstrukcji bardziej spójnego systemu.

Bibliografia

Ajao, E. (2024). *The Iimportance of the 'New York Times' AI Copyright Lawsuit*, https://www.techtarget.com/searchenterpriseai/news/366565172/The-importance-of-the-New-York-Times-AI-copyright-lawsuit?_gl=1*1qc8ady*_ga*MTU3NjQ3NDY3MC4xNzA2MDgyMDI2*_ga_TQKE4GS5P9*MTcwNjc3MDk4Mi4yLjEuMTcwNjc3MTA0MC4wLjAuMA (dostęp: 30.12.2024).

Appel, G., Neelbauer, J., Schweidel, D.A. (2023). *Generative AI Has an Intellectual Property Problem*, https://hbr.org/2023/04/generative-ai-has-an-intellectual-property-problem?utm_medium=paidsearch&utm_source=google&utm_campaign=intlcontent_bussoc&utm_term=Non-Brand&tpcc=intlcontent_bussoc&gad_source=1&gclid=EAIaIQobChMI8eOW7MLUggMVBAsGAB3tkgdxEAAYASAAEgKIYvD_BwE (dostęp: 30.12.2024).

Beyond (2023). *How AI and Emerging Technologies Are Changing the Creative Landscape*, https://www.beyond.agency/blog/how-ai-emerging-technologies-are-changing-the-creative-landscape (dostęp: 30.12.2024).

Czajkowska, M, Walasek, N. (2022). *Czy sztuczna inteligencja może być twórcą wynalazku*, https://hrp.pl/blog/decyzja-dabus) (dostęp: 30.12.2024).

De Cremer, D., Morini Bianzino, M., Falk, B. (2023). *How Generative AI Could Disrupt Creative Work*, https://hbr.org/2023/04/how-generative-ai-could-disrupt-creative-work (dostęp: 30.12.2024).

FT (2023). *British Media and Creative Industries Quizzed Over AI Risks to Copyright*, https://www.ft.com/content/76680815-b313-44a7-b12b-a57ebfcb34bd (dostęp: 30.12.2024).

Ghanghash, S. (2023). *The Interplay Between Artificial Intelligence and Intellectual Property Rights*, https://www.linkedin.com/pulse/interplay-between-artificial-intelligence-property-rights-ghanghash (dostęp: 30.12.2024).

Guler, J. (2023). *The AI Revolution: How Technology is Affecting the Media and Creative Industries?*, https://medium.com/@jackguler/the-ai-revolution-how-technology-is-affecting-the-media-and-creative-industries-d3f924f3d52a (dostęp: 30.12.2024).

Hayden, E. (2023). *Artificial: A Study on the Use of Artificial Intelligence in Art*, https://digitalcommons.unomaha.edu/university_honors_program/208/ (dostęp: 30.12.2024).

Hodge, N. (2023). *AI: Creative Industries Grapple with Technology's Implications for Their IP*, https://www.ibanet.org/AI-creative-industries-grapple-with-tech-implications-for-IP (dostęp: 30.12.2024).

IBM (2023). *The Quest for AI Creativity*, https://www.ibm.com/watson/advantage-reports/future-of-artificial-intelligence/ai-creativity.html (dostęp: 30.12.2024).

Lemley, M.A. (2023). *How Generative AI Turns Copyright Upside Down*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4517702 (dostęp: 30.12.2024).

Logan, S. (2023). *The Impact of AI in Creative Industry*, https://www.twine.net/blog/impact-of-ai-in-creative-industry/ (dostęp: 30.12.2024).

Matulionyte, R. (2022). AI as an Inventor: Has the Federal Court of Australia Erred in DABUS?, *Journal of Intellectual Property, Information Technology and Electronic Commerce Law*, 13(2), https://www.jipitec.eu/jipitec/article/view/348 (dostęp: 30.12.2024).

Neale, E. (2023). *AI NFT: How AI Is Impacting the NFT Scene*, https://nftevening.com/ai-nft-how-ai-is-impacting-the-nft-scene/ (dostęp 30.12.2024).

Ohio (2023). *How AI Is Transforming the Creative Economy and Music Industry*, https://www.ohio.edu/news/2024/01/how-ai-transforming-creative-economy-and-music-industry (dostęp: 30.12.2024).

Saran, C. (2024). *Legal Cases Question IP in Large Language Model Training*, https://www.computerweekly.com/news/366566495/Legal-cases-question-IP-in-Large-Language-Model-training?_gl=1*1xyp1k6*_ga*MTU3NjQ3NDY3MC4xNzA2MDgyMDI2*_ga_TQKE4GS5P9*MTcwNjc3MDk4Mi4yLjEuMTcwNjc3MTI3Ni4wLjAuMA (dostęp: 30.12.2024).

Sperling, J. (2023). *Beyond the Binary: Rethinking the Role of AI in Creative Industries*, https://leading.business.columbia.edu/main-pillar-digital-future/digital-future/ai-creative-industries (dostęp: 30.12.2024).

Tesseract (2023). *Artificial Intelligence in the Creative Industries*, https://tesseract.academy/artificial-intelligence-in-the-creative-industries/ (dostęp: 30.12.2024).

15.0

WPŁYW WDROŻENIA AI NA POSTRZEGANIE METAVERSE W SEKTORZE PRZEMYSŁÓW KREATYWNYCH

Aneta Siejka
Szkoła Główna Handlowa w Warszawie
Rafał Kasprzak
Szkoła Główna Handlowa w Warszawie

Streszczenie

Sztuczna inteligencja (AI) i metaverse to innowacyjne technologie, które mają potencjał znacząco przekształcić procesy wielu organizacji. Chociaż AI jest coraz częściej wykorzystywana, to metaverse wciąż pozostaje w fazie rozwoju. W opisanym badaniu ilościowym ustalono wpływ wdrożenia AI na postrzeganie metaverse w organizacjach w sektorze przemysłów kreatywnych. Wyniki skupiają się na ocenie różnych aspektów metaverse, takich jak: sprzedaż produktów, praca zdalna, spotkania firmowe, szkolenia i kursy, burze mózgów, konferencje, targi i wystawy oraz onboarding nowych pracowników. Sprawdzono, czy istnieje związek pomiędzy ogólną oceną zastosowań metaverse a dostępnością technologii w obrębie organizacji, które wprowadziły AI. Rozdział stanowi punkt wyjścia do badań nad relacjami między AI i metaverse w organizacjach, a także nad ich wpływem na przyszłość pracy. Badanie dostarcza informacji dla pracowników organizacji rozważających wdrożenie AI i metaverse.

Słowa kluczowe: AI, metaverse, praca zdalna, sprzedaż, szkolenia, kursy

Wprowadzenie

Postępująca integracja sztucznej inteligencji (AI) z metaverse to istotny kierunek rozwoju, który ma potencjał zwiększyć możliwości zastosowania nowych technologii przez wiele organizacji [Bordegoni, Ferrise, 2023; Huynh-The *et al.*, 2022; Soliman *et al.*, 2024]. Każdy etap ewolucji technologicznej, od wczesnych eksperymentów z rzeczywistością wirtualną aż po rozwój sztucznej inteligencji, stanowi fundament dla rozwoju metaverse [Ball, 2022; Dwivedi *et al.*, 2022; Qin, Hui, 2023]. Choć koncepcja metaverse w kontekście zastosowań biznesowych wciąż jest na wstępnym etapie rozwoju, może zrealizować swój potencjał dzięki wdrożeniu AI. Pomimo że ten obszar cieszy się coraz większym zainteresowaniem badaczy, nadal brakuje badań nad długoterminowym wpływem AI na postrzeganie metaverse w różnych organizacjach. Na potrzeby badania zdefiniowano „postrzeganie metaverse" jako zespół przekonań, postaw i oczekiwań pracownika organizacji wobec metaverse, dotyczący zarówno jego natury (wirtualny świat, platforma technologiczna), potencjału (możliwości zastosowania), jak i wyzwań (ryzyka, ograniczenia). Wypełnia ono lukę w badaniach nad długoterminowym wpływem sztucznej inteligencji na postrzeganie metaverse poprzez porównanie organizacji, które wprowadziły AI, z organizacjami, które nie wykorzystały AI pod względem oceny poszczególnych zastosowań metaverse. Objęto nim organizacje w sektorze przemysłów kreatywnych, który na potrzeby dalszych analiz zostanie zdefiniowany jako sfera usług społecznych, obejmująca obszar aktywności gospodarczej ukierunkowanej na tworzenie i komercjalizację produktów kultury, obejmującej różne formy organizacyjne prowadzenia działalności gospodarczej w następujących branżach [Kasprzak, 2013]:

- podsektor sztuk i rzemiosł, obejmujący: sztuki wizualne, sztuki performatywne oraz dziedzictwo narodowe, biblioteki, archiwa;
- podsektor produkcji kreatywnej, obejmujący: programowanie, działalność wydawniczą, produkcję filmową i telewizyjną oraz produkcję radiową i muzyczną;
- podsektor usług kreatywnych, obejmujący: modę i wzornictwo, reklamę i działalność pokrewną oraz architekturę i projektowanie wnętrz.

15.1. Sztuczna inteligencja (AI) w metaverse

Metaverse i sztuczna inteligencja to dwie szybko rozwijające się technologie, które mogą zrewolucjonizować sposób, w jaki żyjemy, pracujemy i wchodzimy ze sobą w interakcje [Soliman *et al.*, 2024]. Integracja sztucznej inteligencji z metaverse jest

niezbędna, aby umożliwić ludziom poruszanie się po światach wirtualnych i fizycznych oraz przetwarzać ogromne ilości danych generowanych w metaverse [Huynh-The *et al.*, 2022]. Technologie VR, AR i MR[1] oferują wyjątkowe możliwości immersyjnych i angażujących doświadczeń w metaverse, a sztuczna inteligencja może być wykorzystana do poprawy tych doświadczeń poprzez spersonalizowane rekomendacje treści, adaptacyjne interfejsy oraz bardziej naturalne i intuicyjne interakcje [Huynh-The *et al.*, 2022; Zhao *et al.*, 2022]. Sztuczna inteligencja odgrywa ważną rolę w rozwoju metaverse, ponieważ jest używana w wielu technologiach wykorzystywanych do budowania światów metaverse i zapewniania użytkownikom wciągających doświadczeń [Soliman *et al.*, 2024]. AI może naśladować funkcje poznawcze związane z ludzką inteligencją, takie jak: rozumienie i reagowanie na język mówiony lub pisany [Moradi Shekofteh, 2023], analizowanie danych [Salah *et al.*, 2023], wydawanie poleceń [Pamucar *et al.*, 2023], aby metaverse mógł działać według zasad określonych przez twórcę [Hwang, Chien, 2022]. Część potencjalnych ryzyk, które wynikają ze stosowania sztucznej inteligencji, wiąże się z utratą pracy, zagrożeniami dla prywatności, dezinformacją, zbytnim poleganiem na technologii [Przegalińska, Jemielniak, 2023].

Usługi edukacyjne i szkoleniowe przechodzą znaczące transformacje w wyniku współczesnych zastosowań AI [Upadhyay, Khandelwal, 2019], które pomagają w ocenie wymagań szkoleniowych, projektowaniu dostosowanych materiałów i dostarczaniu szkoleń w dogodnym dla pracowników miejscu [Maity, 2019]. Sztuczna inteligencja wdrażana w procesach szkoleniowych pracowników może znacząco usprawnić ich efektywność. Algorytmy umożliwiają tworzenie kontekstowych, spersonalizowanych i ukierunkowanych programów szkoleniowych, dostosowanych do indywidualnych potrzeb i predyspozycji każdego pracownika. Dzięki temu szkolenia stają się bardziej efektywne, prowadząc do wymiernej poprawy wydajności i umiejętności pracowników w organizacji. Badania Kasprzaka i Ziółkowskiej [2021] pokazały, że większość respondentów (56%) preferuje pracę lub uczenie się w formie online, a nie w biurze/szkole (40%). W ciągu ostatnich dwóch dekad miejsce pracy ewoluowało

1 W literaturze istnieje wiele różnych pojęć „rzeczywistości":
- Rzeczywistość wirtualna (VR) określa trójwymiarowe środowisko wirtualne stworzone komputerowo, mające na celu odwzorowanie rzeczywistości lub fantastycznego świata.
- Rzeczywistość mieszana (MR), nazywana także rzeczywistością hybrydową, to połączenie elementów świata wirtualnego i fizycznego w celu utworzenia nowego środowiska, które umożliwia interakcje między obiektami wirtualnymi i fizycznymi w czasie rzeczywistym.
- Rzeczywistość rozszerzona (AR) to interaktywne przeżycie rzeczywistego środowiska, w którym obiekty fizyczne są wzbogacane o generowane komputerowo informacje percepcyjne.

Próba uporządkowania różnych pojęć rzeczywistości VR, MR, AR do przedpola metaverse została opisana przez Szpringera [2023].

w kierunku hybrydowych modeli, a metaverse może przyspieszyć ten trend, umożliwiając bezproblemową integrację pracy zdalnej i stacjonarnej.

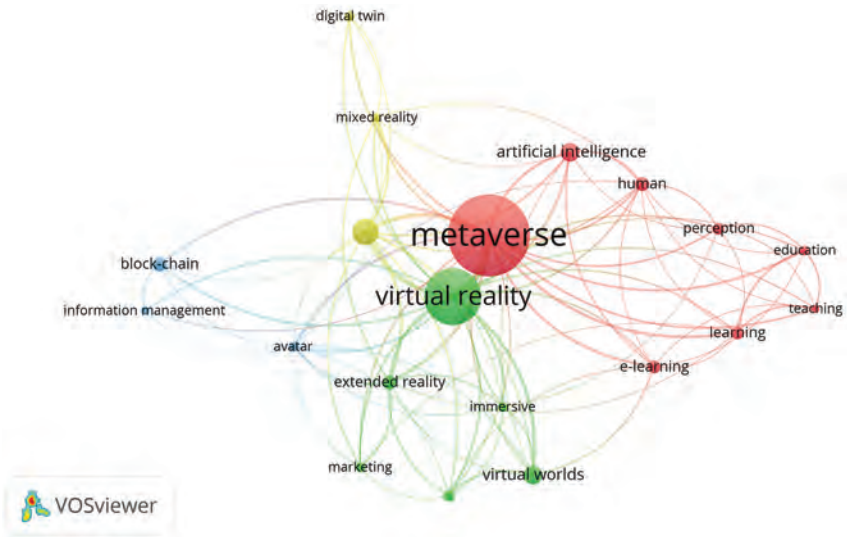
Implementacja sztucznej inteligencji w procesach zarządzania talentami może przyczynić się do zmaksymalizowania potencjału kadrowego organizacji poprzez tworzenie środowiska pracy, dostosowanego do potrzeb pracowników i sprzyjającego produktywności. Pomimo rosnącego trendu nadal wielu przywódców w organizacjach napotyka trudności w dokładnym mapowaniu wpływu sztucznej inteligencji na pracowników oraz na szybkość automatyzacji sektora, w którym działają [Przegalińska, Jemielniak, 2023]. Dlatego przyszłość pracy w kontekście AI wymaga od organizacji ciągłej adaptacji i innowacji. Warto podkreślić, że w gospodarce cyfrowej najnowocześniejsze technologie, takie jak m.in.: sztuczna inteligencja, cyfrowe bliźniaki, rozszerzona rzeczywistość, zostaną opracowane w ramach przemysłu 4.0 jako fundament realizacji metaverse [Mourtzis *et al*., 2022]. Dzięki postępowi technologicznemu, który umożliwia wykorzystanie zaawansowanych rozwiązań współpracy w czasie rzeczywistym, środowisko metaverse ze wsparciem AI może dalej się rozwijać.

15.2. Przegląd istniejących badań

Należy podkreślić, że zagadnienie integracji sztucznej inteligencji z metaverse jest stosunkowo nowe. Analiza publikacji z bazy Scopus z ostatnich trzech lat jasno pokazuje rosnące zainteresowanie badaczy i naukowców tym obszarem. Najwięcej publikacji ukazało się w ostatnich dwóch latach: 2023 (33) i 2022 (14), co wskazuje na zwiększone zainteresowanie łączeniem tych dwóch obszarów badawczych. W ramach pogłębionych badań literaturowych przygotowano wizualizacje mapy słów kluczowych z wykorzystaniem otwartego programu VOSviewer (www.vosviewer.com), służącego do konstruowania i wizualizacji danych bibliometrycznych. Narzędzie posłużyło do graficznego przedstawienia sieci współwystępowania ważnych terminów związanych ze sztuczną inteligencją i metaverse. Za podstawę naukowego warsztatu przyjęto metodę bibliometryczną, w tym analizę współwystępowania słów kluczowych. Źródłem danych do wizualizacji była międzynarodowa interdyscyplinarna baza Scopus, która zawiera informacje o opublikowanych pracach naukowych, takich jak artykuły w czasopismach naukowych w języku angielskim, ograniczone do obszarów tematycznych z nauk społecznych oraz biznesu, zarządzania i rachunkowości. W bazie Scopus zidentyfikowano 63 artykuły (stan z lipca 2024 r.), w których tytule, abstrakcie lub słowach kluczowych występują pojęcia: „sztuczna inteligencja" lub „metaverse". Wizualizacją objęto artykuły opublikowane w czasopismach indek-

sowanych. Program jest oparty na technice VOS (*visulatization of similarities*) i służy do wizualizacji podobieństw między obiektami, w której obiekty podobne są zlokalizowane blisko siebie, a obiekty mniej podobne są od siebie oddalone [van Eck, Waltman, 2007]. Zjawisko oddalenia obiektów dobrze ilustruje przykład zamieszczony na rysunku 15.1.

Rysunek 15.1. Mapa słów kluczowych – oddalenie obiektów



Źródło: opracowanie własne przy użyciu programu VOSviewer.

Na szczególną uwagę zasługuje grupa słów w węźle czerwonym, związanych z percepcją (*perception*), edukacją (*education*), nauczaniem *(teaching*), nauką (*learning*) oraz e-learningiem, co wskazuje na zwiększone możliwości dotyczące obszarów związanych z nauczaniem w środowisku metaverse, napędzanym przez sztuczną inteligencję. Grupa słów oznaczona kolorem zielonym nawiązuje głównie do właściwości technologicznych, takich jak: rozszerzona rzeczywistość (*extended reality*), immersyjność (*immersive*), wirtualny świat (*virtual world*), ale również podkreśla powiązania z marketingiem, a tym samym tworzenie kampanii marketingowych, które wykorzystują te technologie do dotarcia do nowych odbiorców i zwiększania sprzedaży. Trzeci węzeł w kolorze niebieskim obejmuje słowa: avatar, zarządzanie informacją (*information management*), blockchain, co oznacza zintegrowanie tych elementów w celu stworzenia nowego, zdecentralizowanego systemu zarządzania tożsamością i danymi. Ostatnia żółta grupa słów kluczowych zawiera: mieszaną rzeczywistość (*mixed reality*) oraz cyfrowe bliźniaki (*digital twins*), które łączą się ze

sztuczną inteligencją (*artificial intelligence*), co oznacza, że połączenie tych technologii może umożliwić tworzenie interaktywnych i realistycznych wizualizacji, które mogą być wykorzystywane do edukacji i nauczania. Mapa słów kluczowych wskazuje na szeroki zakres możliwości związanych z metaverse, napędzanym sztuczną inteligencją w różnych dziedzinach, w tym w edukacji, marketingu i zarządzaniu danymi. Naukowcy i praktycy eksplorują zastosowania sztucznej inteligencji w metaverse, co z kolei może przyczynić się do dalszego rozwoju i stworzenia bardziej wciągającego i interaktywnego wirtualnego świata.

15.3. Charakterystyka środowiska metaverse

Metaverse można opisać jako połączenie światów fizycznego i cyfrowego. Wizja metaverse zakłada powstanie trójwymiarowego, wirtualnego świata zintegrowanego z Internetem, w którym użytkownicy będą mogli w czasie rzeczywistym wchodzić w interakcje ze sobą i z obiektami cyfrowymi. Ta koncepcja jest skomplikowana i często omawiana jako wielowymiarowa, ponieważ zmienia się wraz z postępem technologicznym. Ze względu na swoją złożoność nie ma obecnie jasnego konsensusu co do tego, jak należy definiować lub opisywać metaverse. Kluczowy problem, wpływający na wykształcenie się wielu definicji i braku spójnego stanowiska wśród różnych badaczy, wynika z faktu, że choć ten cyfrowy wszechświat już istnieje, wciąż jest w fazie intensywnej rozbudowy[2]. Istnieje bardzo wiele ujęć znaczenia terminu „metaverse". To wizja trójwymiarowego wirtualnego świata, w którym awatary angażują się w działania polityczne, gospodarcze, społeczne i kulturalne [Park, Kim, 2022]. Według słownika Oxford English [2024] termin „metaverse" składa się etymologicznie z dwóch elementów: meta (grecki przedrostek *μετα* oznaczający post, po lub poza – oznaczający zmianę, transformację, permutację lub substytucję) i wszechświat. W literaturze funkcjonuje wiele określeń i definicji metaverse, które są różnorodnie pojmowane w zależności od kontekstu. Pojawiają się wąskie definicje, odnoszące się do konkretnych aspektów, takich jak interaktywne światy wirtualne czy autonomiczni agenci [Singla *et al*., 2023]. Ponadto istnieją też bardziej rozbudowane koncepcje, które opisują metaverse jako – immersyjną wersję Internetu [Lim *et al*., 2022], pod uwagę brane

[2] Nawet bardzo skrótowa próba przedstawienia różnorodnego dorobku literatury i badań związanych z technologicznymi aspektami metaverse wykracza poza zakres tego rozdziału. Warto jednak wspomnieć, że zbudowanie metaverse wymagać będzie infrastruktury sieciowej i obliczeniowej. Zagadnienia związane z silnikiem gier i platformami obsługującymi wiele wirtualnych światów, standardami niezbędnymi do ich integracji, narzędzi dostępu oraz systemy to tematy omówione przez Balla [2022].

są także doświadczenia AR, VR i MR [Nevelsteen, 2018; Park *et al.*, 2022; Rauschnabel *et al.*, 2022; Hwang, Chien, 2022]. Pojawia się również pojęcie „ekosystemów metaverse", po których można poruszać się pod kontrolą użytkownika [Lim *et al.*, 2023], oraz związane z nimi ryzyka, dotyczące m.in. legislacji, własności, kontroli, oszustw, zagrożeń prywatności, etyki, odpowiedzialności [Smaili *et al.*, 2022].

Warto podkreślić, że temat adaptacji i regulacji metaverse został uwzględniony w agendzie prac Unii Europejskiej jako podstawa do ustrukturyzowania ekosystemu oraz jako plan rozwoju dotyczący wirtualnych światów, takich jak metaverse [European Commission – Statement, 2022]. Będzie to miało wpływ na popyt, mechanizmy cenowe oraz pojawianie się wirtualnych miejsc pracy i źródeł dochodu [Suganya, Sasirekha, 2024]. Od lat przedsiębiorstwa badają nowe metody komunikacji z klientami, pracownikami i partnerami za pośrednictwem metaverse, aby uzyskać przewagę konkurencyjną, zwiększyć zaangażowanie i generować nowe strumienie przychodów [Nouman *et al.*, 2023].

W gospodarce cyfrowej wraz z rozwojem pracy zdalnej tradycyjne spotkania osobiste coraz częściej zastępowane są ich wirtualnymi odpowiednikami. Przedsiębiorstwa wykorzystują wirtualne przestrzenie spotkań w środowisku metaverse do zdalnej współpracy i komunikacji. Umożliwia to pracownikom z różnych lokalizacji interakcję w czasie rzeczywistym przy jednoczesnym pokonywaniu ograniczeń geograficznych. W metaverse jednostki mogą wchodzić w interakcje z innymi osobami z różnych lokalizacji geograficznych, tak jakby byli razem fizycznie obecni [Bale *et al.*, 2022]. Metaverse stwarza również nowe perspektywy dla pracy i edukacji. Platforma ta oferuje innowacyjne rozwiązania w zakresie współpracy i kształcenia, co wpisuje się w rosnącą popularność pracy zdalnej i nauki online. Dzięki narzędziom do wirtualnych tablic i prezentacji wirtualne biura i klasy mogą zapewnić bardziej angażujące i partycypacyjne doświadczenie [Krotoski, 2022]. Może to ułatwić tworzenie zdalnych przestrzeni roboczych, umożliwiając ludziom interakcję i współpracę w tym środowisku. W konsekwencji zmniejszy to bariery geograficzne i zapewni zespołom rozproszonym efektywną współpracę. Technologia VR może zostać wykorzystana do stworzenia immersyjnych doświadczeń we wdrażaniu nowych pracowników. W wirtualnym otoczeniu te osoby mogą poznać firmę, kulturę organizacyjną i swoje obowiązki w bardziej angażujący i interaktywny sposób, co może przyspieszyć proces wdrażania i poprawić retencję pracowników. Choć koncepcja metaverse w kontekście zastosowań biznesowych wciąż jest na wstępnym etapie rozwoju, ma duży potencjał zmiany w organizacjach.

15.4. Zastosowania metaverse i AI w organizacji – wyniki badań

Celem badania jest analiza wpływu wdrożenia sztucznej inteligencji (AI) na postrzeganie metaverse w organizacji. Wyniki skupiają się na ocenie różnych aspektów metaverse, takich jak: sprzedaż produktów, praca zdalna, spotkania firmowe, szkolenia i kursy, burze mózgów, konferencje, targi i wystawy oraz onboarding nowych pracowników. Przeprowadzono badanie ankietowe metodą CATI (*computer-assisted telephone interviewing*, czyli wspomagany komputerowo wywiad telefoniczny) w okresie 21.08–15.11.2023 r. wśród 141 pracowników organizacji kreatywnych, jako podmiotów działających w sektorze przemysłów kreatywnych (tabela 15.1), mających potencjał realizacji swoich celów w środowisku metaverse.

Tabela 15.1. Sektor przemysłów kreatywnych objęty badaniem

	Podsektor	Branże i kody PKD	Organizacje, które wykorzystały AI (N = 60)	Organizacje, które nie wykorzystywały AI (N = 81)
Sektor przemysłów kreatywnych	Podsektor sztuk i rzemiosł	sztuki wizualne (74.20.Z, 90.03.Z i 47.78.Z)	3	4
		sztuki performatywne (90.01.Z, 90.02.Z i 90.04.Z)	1	7
		dziedzictwo narodowe, biblioteki archiwa (91.01.A, 91.01.B i 91.02.Z)	5	13
	Podsektor sztuk i rzemiosł	programowanie (91.01.A, 91.01.B i 91.02.Z)	15	6
		działalność wydawnicza (58.11.Z, 58.13. Z, 58.14.Z i 58.19.Z)	6	7
		produkcja filmowa i telewizyjna (59.11.Z, 59.13.Z i 59.14.Z)	1	3
		produkcja radiowa i muzyczna (59.20.Z, 60.10.Z i 60.20.Z)	2	3
	Podsektor usług kreatywnych	moda i wzornictwo (74.10.Z)	3	9
		reklama i działalność pokrewna (73.11.Z, 73.12.A, 73.12.B, 73.12.C i 73.12.D)	14	20
		architektura i projektowanie wnętrz (71.11.Z)	10	9
RAZEM			60	81

Źródło: opracowanie na podstawie: Kasprzak [2013].

Badanie naukowe zrealizowane zostało w ramach „Grantu Dziekana Szkoły Doktorskiej SGH". Uwzględniono w nim zarówno organizacje, które wykorzystały AI (N = 60), jak i te, które tego nie zrobiły (N = 81). Pierwszym etapem analizy było

porównanie poszczególnych aspektów zastosowań metaverse w obrębie organizacji, które wprowadziły AI. W tym celu wykonano test Friedmana, a jego wyniki przedstawiono w tabeli 15.2.

Tabela 15.2. Porównanie oceny poszczególnych aspektów zastosowań metaverse w obrębie organizacji, które wcześniej wprowadziły AI (*N* = 60)

W jakich obszarach widzi Pan/i zastosowanie metaverse?	Średnia ranga	*M*	*SD*	$\chi^2(7)$	*P*	*W*
Sprzedaż produktów	4,55	3,80	1,09	19,57	**0,007**	0,07
Praca zdalna	4,10	3,62	1,04			
Spotkania firmowe	4,15	3,53	1,10			
Szkolenia i kursy	5,27	4,03	0,94			
Burze mózgów	4,79	3,95	0,67			
Konferencje	4,83	3,87	1,03			
Targi i wystawy	4,03	3,58	1,05			
Onboarding nowych pracowników	4,28	3,67	1,00			

Uwagi: *N* – liczba obserwacji, *M* – średnia, *SD* – odchylenie standardowe, χ^2 – wartość statystyki testowej, *p* – istotność statystyczna, *W* – wskaźnik siły efektu.

Źródło: opracowanie własne.

Analiza wykazała, że w obrębie organizacji, które wcześniej wprowadziły AI, występowała istotna statystycznie różnica w zakresie oceny poszczególnych aspektów zastosowań metaverse. Okazało się jednak, że odnotowany efekt okazał się słaby ($W \leq 0{,}10$). W celu zbadania istoty różnic przeprowadzono test *post-hoc* Dunn, którego wyniki przedstawiono w tabeli 15.3.

Tabela 15.3. Wyniki testu post-hoc dla porównania oceny poszczególnych aspektów zastosowań metaverse w obrębie organizacji, które wcześniej wprowadziły AI (*N* = 60)

	Sprzedaż produktów	Praca zdalna	Spotkania firmowe	Szkolenia i kursy	Burze mózgów	Konferencje	Targi i wystawy	Onboarding nowych pracowników
Sprzedaż produktów	–							
Praca zdalna	1,01	–						
Spotkania firmowe	0,89	–0,11	–					
Szkolenia i kursy	–1,60	–2,61**	–2,50*	–				

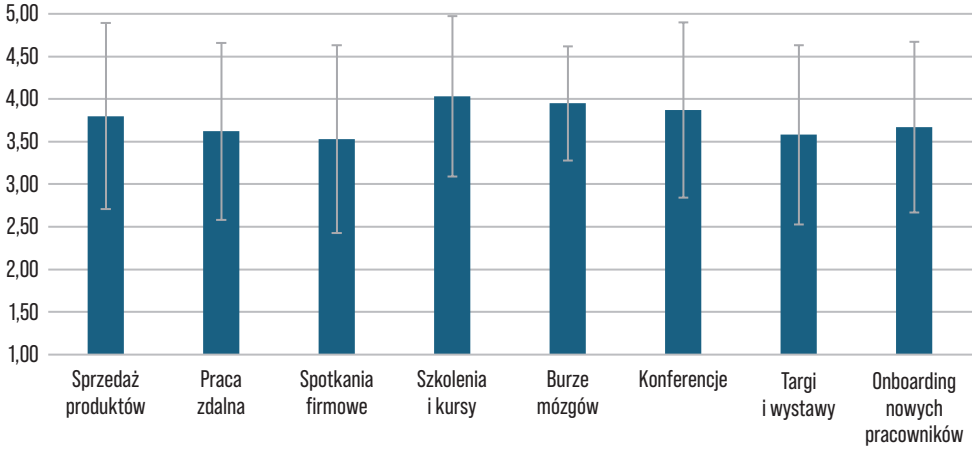
cd. tabeli 15.3

	Sprzedaż produktów	Praca zdalna	Spotkania firmowe	Szkolenia i kursy	Burze mózgów	Konferencje	Targi i wystawy	Onboarding nowych pracowników
Burze mózgów	–0,54	–1,55	–1,43	1,06	–			
Konferencje	–0,61	–1,62	–1,51	0,99	–0,07	–		
Targi i wystawy	1,16	0,15	0,26	2,76**	1,70	1,77	–	
Onboarding nowych pracowników	0,60	–0,41	–0,30	2,20*	1,14	1,21	–0,56	–

*** $p < 0{,}001$; ** $p < 0{,}010$; * $p < 0{,}050$.
Źródło: opracowanie własne.

Otrzymane wyniki wskazywały na to, że pracownicy organizacji, które wcześniej wprowadziły AI, oceniali zastosowanie metaverse w zakresie szkoleń i kursów istotnie wyżej niż w zakresie pracy zdalnej, spotkań firmowych, targów i wystaw oraz onboardingu nowych pracowników. Rezultaty analizy zobrazowano na rysunku 15.2.

Rysunek 15.2. Wartości średnie wraz z odchyleniami standardowymi oceny poszczególnych aspektów zastosowań metaverse w obrębie organizacji, które wcześniej wprowadziły AI (pytanie: „W jakich obszarach widzi pan/i zastosowanie metaverse?")

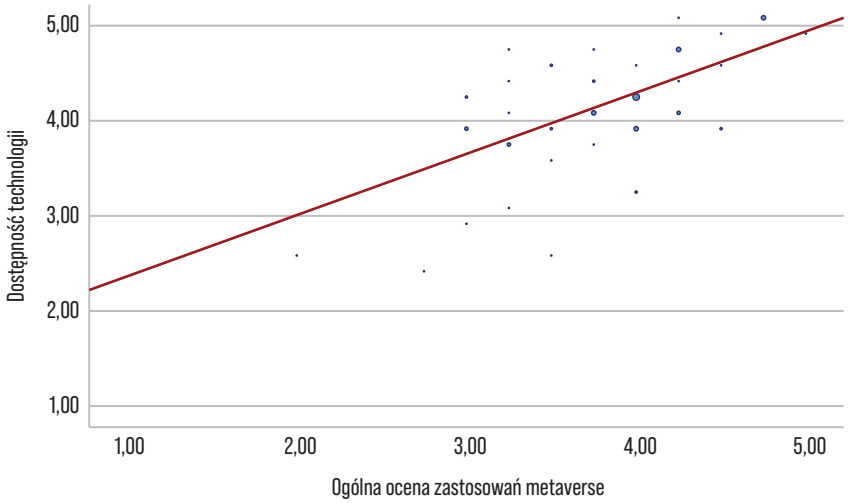


Źródło: opracowanie własne.

Następnie sprawdzono, czy istnieje związek pomiędzy ogólną oceną zastosowań metaverse a dostępnością technologii w obrębie organizacji, które wcześniej wprowadziły AI. W tym celu wykonano analizę korelacji *r* Pearsona. Wykazała ona, że w gru-

pie pracowników organizacji, które wprowadziły AI, występował istotny statystycznie dodatni związek ($r = 0{,}63$; $p < 0{,}001$) pomiędzy ogólną oceną zastosowań metaverse a oceną dostępnością technologii. Oznacza to, że wraz ze wzrostem oceny zastosowań metaverse rosła ocena dostępności technologicznej. Wyniki analizy zobrazowano na rysunku 15.3.

Rysunek 15.3. Wykres rozrzutu wraz z linią dopasowania dla związku pomiędzy ogólną oceną zastosowań metaverse a dostępnością technologii w obrębie organizacji, które wprowadziły AI



Źródło: opracowanie własne.

W dalszej kolejności sprawdzono, czy organizacje, które wprowadziły AI, różniły się pod względem oceny poszczególnych zastosowań metaverse od organizacji, które tego nie zrobiły. W tym celu przeprowadzono test Manna-Whitneya, a jego wyniki przedstawiono w tabeli 15.4.

Analiza wykazała istotną statystycznie różnicę pomiędzy porównywanymi grupami jedynie w zakresie oceny zastosowań metaverse w dziedzinach onboardingu nowych pracowników. Okazało się, że pracownicy organizacji, które wprowadziły AI, istotnie wyżej oceniali zastosowanie metaverse we wskazanym obszarze, w porównaniu z pracownikami organizacji, które nie wykorzystywały AI. Należy jednak zwrócić uwagę na fakt, że odnotowany efekt okazał się słaby ($\eta^2 < 0{,}06$). Rezultaty analizy zobrazowano na rysunku 15.4.

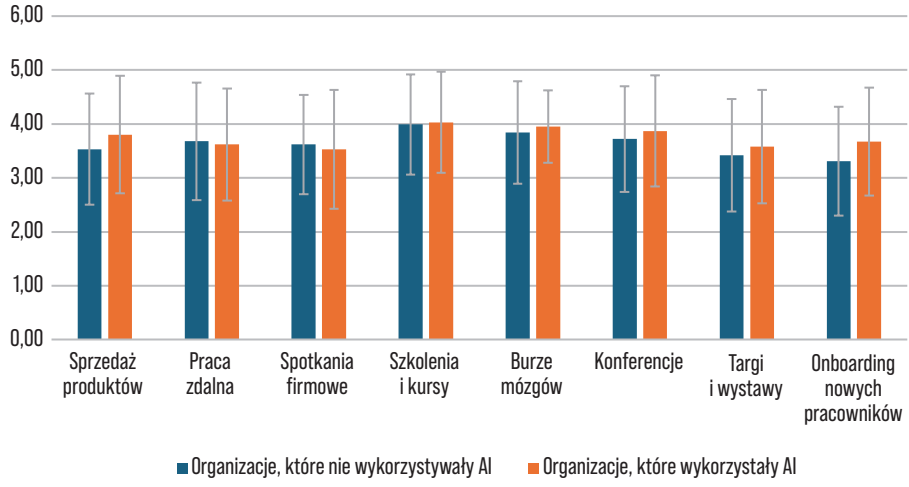
Tabela 15.4. Porównanie organizacji, które wprowadziły AI, z organizacjami, które nie wprowadziły AI, pod względem oceny poszczególnych zastosowań metaverse (*N* = 141)

	Organizacje, które nie wykorzystywały AI (*N* = 81)			Organizacje, które wykorzystały AI (*N* = 60)					
W jakich obszarach widzi pan/i zastosowanie metaverse?	średnia ranga	*M*	*SD*	średnia ranga	*M*	*SD*	*Z*	*P*	η^2
Sprzedaż produktów	66,40	3,53	1,03	77,22	3,80	1,09	–1,64	0,101	0,02
Praca zdalna	72,52	3,68	1,09	68,94	3,62	1,04	–0,54	0,590	<0,01
Spotkania firmowe	72,02	3,62	0,92	69,62	3,53	1,10	–0,37	0,713	<0,01
Szkolenia i kursy	70,19	3,99	0,93	72,09	4,03	0,94	–0,29	0,769	<0,01
Burze mózgów	70,34	3,84	0,95	71,89	3,95	0,67	–0,25	0,806	<0,01
Konferencje	67,94	3,72	0,98	75,13	3,87	1,03	–1,09	0,278	<0,01
Targi i wystawy	68,49	3,42	1,04	74,39	3,58	1,05	–0,89	0,374	<0,01
Oboarding nowych pracowników	64,79	3,31	1,01	79,38	3,67	1,00	–2,19	**0,029**	0,03

Uwagi: *N* – liczba obserwacji, *M* – średnia, *SD* – odchylenie standardowe, *Z* – wartość statystyki testowej, *p* – istotność statystyczna, η^2 – wskaźnik siły efektu.

Źródło: opracowanie własne.

Rysunek 15.4. Wartości średnie wraz z odchyleniami standardowymi oceny poszczególnych obszarów zastosowania metaverse wśród organizacji, które wprowadziły lub nie wprowadziły AI (pytanie: „W jakich obszarach widzi pan/i zastosowanie metaverse?")



Źródło: opracowanie własne.

15.5. Omówienie wyników badania

Wyniki badania sugerują, że wdrożenie AI może kształtować pozytywne postrzeganie metaverse w organizacji w zakresie wielu zastosowań. Jest to spójne z poglądem innych badaczy, którzy udowadniają, że generatywna sztuczna inteligencja może wyposażyć metaverse w szeroką gamę zastosowań, w tym sprzedaż produktów, pracę zdalną i szkolenia [Qin, Hui, 2023]. Ponadto wraz ze wzrostem oceny zastosowań metaverse rosła ocena dostępności technologicznej w grupie pracowników przedsiębiorstw z sektora przemysłów kreatywnych, które wprowadziły AI. Jest to zbieżne z innymi badaniami, które również podkreślają, że w przypadku organizacji wdrażających zasady AI skuteczne przyjęcie zależy od czynników, takich jak komunikacja, wsparcie kierownictwa, szkolenia i infrastruktura techniczna [Kelley, 2022]. Badanie podkreśliło znaczenie szkoleń i kursów, co związane jest z uczeniem się w organizacji – potwierdzają to inne badania, świadczące o tym, że sztuczna inteligencja wykazała swoją znaczącą wartość dla tych technologii, szczególnie w związku z powstawaniem i rozwojem głębokiego uczenia się [Guo *et al.*, 2022]. Co więcej, trening w metaverse również mógłby zastąpić lub przynajmniej udoskonalić szkolenie online, jeśli weźmie się pod uwagę jego możliwości monitorowania np. postępów w szkoleniach i przetwarzania wyników szkoleń (tj. gry biznesowe, testy) poprzez wykorzystanie możliwości, jakie daje sztuczna inteligencja [Marabelli, Lirio, 2024]. Niektóre badania omawiają, w jaki sposób metaverse, *digital twins* i AI mogą usprawnić proces zdobywania wiedzy, umiejętności i *know-how*, ze szczególnym uwzględnieniem sektora przemysłowego [Bordegoni, Ferrise, 2023]. Istnieją także opracowania skupiające się na sposobie, w jaki metaverse można wykorzystać do rozwoju, prototypowania, oceny, regulacji i udoskonalania praktyki medycznej opartej na sztucznej inteligencji [Wang *et al.*, 2022]. Inni badacze przedstawiają zaś perspektywy zastosowania metaverse w kilku ważnych obszarach, tj. marketingu, turystyki, produkcji, zarządzania operacyjnego, edukacji, branży detalicznej, usług bankowych, opieki zdrowotnej i zarządzania zasobami ludzkimi [Koohang *et al.*, 2023].

A zatem potrzeba dalszych badań nad wpływem AI na postrzeganie i akceptację metaverse. Opisane badanie ma pewne ograniczenia, ponieważ zostało przeprowadzone na próbie pracowników organizacji wyłącznie z sektora kreatywnego. Wyniki mogą nie dotyczyć innych sektorów gospodarki. Badanie nie uwzględniło potencjalnych negatywnych skutków integracji AI z metaverse, co wymaga dalszych badań.

15.6. Zastosowania metaverse i AI w organizacji – kierunki dalszych badań

Sztuczna inteligencja (AI) posiada potencjał znaczącego wzbogacenia doświadczeń pracowników w metaverse oraz usprawnienia ich funkcjonowania w wirtualnym środowisku organizacji. Z tego względu kluczowe znaczenie ma kontynuowanie badań i rozwoju tych technologii, a także integracja w sposób maksymalizujący ich potencjału pozytywnego oddziaływania. Przeprowadzone badanie stanowi punkt wyjścia do dalszych badań nad wpływem sztucznej inteligencji na postrzeganie i akceptację metaverse w różnych organizacjach. Warto również uwzględnić przyszłe kierunki badań, aby lepiej zrozumieć potencjał metaverse, napędzanego sztuczną inteligencją i jego wpływu na organizacje, co ilustruje tabela 15.5.

Tabela 15.5. Rozwój metaverse poprzez integrację ze sztuczną inteligencją – przyszłe kierunki badań

Kategoria	Obszar badawczy	Opis
1. Strategie wdrożenia AI	określenie sposobów integracji AI z metaverse dla osiągnięcia strategicznych celów organizacyjnych	opracowanie modeli wdrożenia dostosowanych do różnych branż
2. Mierzalne korzyści	analizowanie wymiernych korzyści płynących z wdrożenia AI w metaverse	szukanie metod oceny ROI (zwrotu z inwestycji) dla projektów AI w metaverse
3. Bezpieczeństwo	zmapowanie negatywnych kwestii związanych z prywatnością i bezpieczeństwem danych w metaverse napędzanym AI	opracowanie rozwiązań zapewniających bezpieczeństwo danych w metaverse
4. Kwestie etyczne i prawne	badanie potencjalnych zagrożeń etycznych związanych z wykorzystaniem AI w metaverse	sporządzenie wytycznych etycznych dla odpowiedzialnego rozwoju i stosowania AI w metaverse
5. Wpływ na kulturę organizacyjną	uwzględnienie wpływu AI na sposób interakcji i komunikacji pracowników w metaverse	analiza wpływu AI na strukturę organizacyjną i procesy decyzyjne

Źródło: opracowanie własne.

Integracja technologii sztucznej inteligencji z metaverse może z pewnością rozszerzyć istniejące procesy biznesowe, otwierając nowe możliwości i kanały działania. Aby jednak odnieść sukces w tym dynamicznym środowisku, organizacje muszą rozważyć szereg kluczowych aspektów. Należy wziąć pod uwagę czynniki, takie jak specyfika branży, wielkość organizacji, zasoby technologiczne i umiejętności pracowników.

Podsumowanie

Sztuczna inteligencja (AI) i metaverse należą do najszybciej rozwijających się trendów, które mogą zmienić sposób funkcjonowania przedsiębiorstw w sektorze przemysłów kreatywnych. W tym kontekście istotne staje się zrozumienie wpływu wdrożenia AI na postrzeganie metaverse w organizacjach. Badania wskazują na pozytywny wpływ wdrożenia AI na zastosowanie metaverse w zakresie szkoleń i kursów istotnie wyżej niż w zakresie pracy zdalnej, spotkań firmowych, targów i wystaw oraz onboardingu nowych pracowników w sektorze przemysłów kreatywnych. Metaverse może być wykorzystywany do szerokiego spektrum celów, m.in. sprzedaży produktów, pracy zdalnej, spotkań firmowych, szkoleń i kursów, burz mózgów, konferencji, targów i wystaw oraz onboardingu nowych pracowników. Metaverse i AI to stosunkowo nowe technologie, których wpływ na organizacje jest wciąż badany. Potrzebne są dalsze badania, aby lepiej zrozumieć, jak te technologie mogą być wykorzystywane do poprawy wydajności i innowacyjności organizacji.

Bibliografia

Bale, A.S., Ghorpade, N., Hashim, M.F., Vaishnav, J., Almaspoor, Z. (2022). A Comprehensive Study on Metaverse and Its Impacts on Humans, *Advances in Human-Computer Interaction*, s. 1–11. DOI: 10.1155/2022/3247060.

Ball, M. (2022). *Metawersum. Jak internet przyszłości zrewolucjonizuje świat i biznes*. Warszawa: MT Biznes.

Bordegoni, M., Ferrise, F. (2023). Exploring the Intersection of Metaverse, Digital Twins, and Artificial Intelligence in Training and Maintenance, *Journal of Computing and Information Science in Engineering*, 23(6), s. 1–11. DOI: 10.1115/1.4062455.

Dwivedi, Y.K., Hughes, L., Baabdullah, A.M., Ribeiro-Navarrete, S., Giannakis, M., Al-Debei, M.M., Dennehy, D., Metri, B., Buhalis, D., Cheung, C.M.K., Conboy, K., Doyle, R., Dubey, R., Dutot, V., Felix, R., Goyal, D.P., Gustafsson, A., Hinsch, C., Jebabli, I., … Wamba, S.F. (2022). Metaverse Beyond the Hype: Multidisciplinary Perspectives on Emerging Challenges, Opportunities, and Agenda for Research, Practice and Policy, *International Journal of Information Management*, 66, 102542. DOI: 10.1016/j.ijinfomgt.2022.102542.

European Commission – Statement (2022). *People, Technologies & Infrastructure – Europe's Plan to Thrive in the Metaverse I Blog of Commissioner Thierry Breton*, https://ec.europa.eu/commission/presscorner/api/files/document/print/en/statement_22_5525/STATEMENT_22_5525_EN.pdf (dostęp: 8.07.2024).

Guo, Y., Yu, T., Wu, J., Wang, Y., Wan, S., Zheng, J., Fang, L., Dai, Q. (2022). Artificial Intelligence for Metaverse: A Framework, *CAAI Artificial Intelligence Research*.

Hwang, G.J., Chien, S.Y. (2022). Definition, Roles, and Potential Research Issues of the Metaverse in Education: An Artificial Intelligence Perspective, *Computers and Education: Artificial Intelligence*, 3, 100082.

Huynh-The T., Pham Q.-V., Pham X.-Q., Nguyen T.T., Han Z., Kim D.-S. (2022). *Artifcial Intelligence for the Metaverse: A Survey*. DOI: 10.48550/arXiv.2202.10336.

Kasprzak, R. (2013). *Przemysły kreatywne w Polsce: perspektywy i uwarunkowania*. Warszawa: Kamon Consulting.

Kasprzak, R., Ziółkowska, M. (2021). *Badanie: Praca i nauka zdalna w czasach pandemii*, https://gazeta.sgh.waw.pl/meritum/badanie-praca-i-nauka-zdalna-w-czasach-pandemii (dostęp: 8.07.2024).

Kelley, S. (2022). Employee Perceptions of the Effective Adoption of AI Principles, *Journal of Business Ethics*, 178(4), s. 871–893. DOI: 10.1007/s10551-022-05051-y.

Koohang, A., Nord, J.H., Ooi, K., Tan, G., Al-Emran, M., Aw, E.C., Baabdullah, A., Buhalis, D., Cham, T., Dennis, C., Dutot, V., Dwivedi, Y.K., Hughes, L., Mogaji, E., Pandey, N., Phau, I., Raman, R., Sharma, A., Sigala, M., Ueno, A., Wong, L.W. (2023). Shaping the Metaverse into Reality: A Holistic Multidisciplinary Understanding of Opportunities, Challenges, and Avenues for Future Investigation, *Journal of Computational Information Systems*, 63(3), s. 1–31. DOI: 10.1080/08874417.2023.2165197.

Krotoski, A. (2022). *A Beginner's Guide to the Metaverse: What It Is, how You Can Access It and More*, https://www.sciencefocus.com/future-technology/metaverse (dostęp: 8.07.2024).

Lim, W.Y.B., Xiong, Z., Niyato, D., Cao, X., Miao, C., Sun, S., Yang, Q. (2022). Realizing the Metaverse with Edge Intelligence: A Match made in Heaven, *IEEE Wireless Communications*, 30(4), s. 64–71.

Maity, S. (2019), Identifying Opportunities for Artificial Intelligence in the Evolution of Training and Development Practices, *Journal of Management Development*, 38(8), s. 651–663.

Marabelli, M., Lirio, P. (2024). AI and the Metaverse in the Workplace: DEI Opportunities and Challenges, *Personnel Review* (w druku). DOI: 10.1108/PR-04-2023-0300.

Mourtzis, D., Panopoulos, N., Angelopoulos, J., Wang, B., Wang, L. (2022). Human Centric Platforms for Personalized Value Creation in Metaverse, *Journal of Manufacturing Systems*, 65, s. 653–659.

Moradi, A., Shekofteh, Y. (2023). Spoken Language Identifcation Using a Genetic-Based Fusion Approach to Combine Acoustic and Universal Phonetic Results, *Computers and Electrical Engineering*, 105, 108549. DOI: 10.1016/j.compeleceng.2022.108549.

Nevelsteen, K.J. (2018). Virtual World, Defined from a Technological Perspective and Applied to Video Games, Mixed Reality, and the Metaverse, *Computer Animation and Virtual Worlds*, 29(1), s. e1752.

Nouman, S., Lagishetty, S., Reddy, P.P., Chola, C. (2023). Evaluation and Future Impact of a Metaverse in Business. W: *The Business of the Metaverse* (s. 15–30), K. Hemachandran, R.V. Rodriguez (Eds.). New York: Productivity Press.

Oxford English Dictionary (2024). *Metaverse*. DOI: 10.1093/OED/2615256312.

Pamucar, D., Deveci, M., Gokasar, I., Delen, D., Köppen, M., Pedrycz, W. (2023). Evaluation of Metaverse Integration Alternatives of Sharing Economy in Transportation Using Fuzzy Schweizer-Sklar Based Ordinal Priority Approach, *Decis Support Systems*, 171 (C). DOI: 10.1016/j.dss.2023.113944.

Park, S.M., Kim, Y.G. (2022). A Metaverse: Taxonomy, Components, Applications, and Open Challenges, *IEEE Access*, 10, s. 4209–4251. DOI: 10.1109/ACCESS.2021.3140175.

Przegalińska, A., Jemielniak, D. (2023). *AI w strategii: rewolucja sztucznej inteligencji w zarządzaniu*. Warszawa: MT Biznes.

Qin, H.X., Hui, P. (2023). Empowering the Metaverse with Generative AI: Survey and Future Directions. W: *2023 IEEE 43 rd International Conference on Distributed Computing Systems Workshops* (s. 85–90). Hong Kong: IEEE. DOI: 10.1109/ICDCSW60045.2023.00022.

Rauschnabel, P.A., Felix, R., Hinsch, C., Shahab, H., Alt, F. (2022). What is XR? Towards a Framework for Augmented and Virtual Reality, *Computers in Human Behavior*, 133, 107289.

Salah, M., Halbusi, H.A., Abdelfattah, F. (2023). May the Force of Text Data Analysis be with You: Unleashing the Power of Generative AI for Social Psychology Research, *Computers in Human Behavior: Artificial Humans*, 1(2), 100006. DOI: 10.1016/j.chbah.2023.100006.

Singla, A., Gupta, N., Aeron, P., Jain, A., Garg, R., Sharma, D., Gupta, B.B., Arya, V. (2023). Building the Metaverse: Design Considerations, Socio-Technical Elements, and Future Research Directions of Metaverse, *Journal of Global Information Management*, 31(2), s. 1–28.

Smaili, N., de Rancourt-Raymond, A. (2022). Metaverse: Welcome to the New Fraud Marketplace, *Journal of Financial Crime*, 31(1), s. 188–200.

Soliman, M.M., Ahmed, E., Darwish, A., Hassanien, A.E. (2024). Artificial Intelligence powered Metaverse: Analysis, Challenges and Future Perspectives, *Artificial Intelligence Review*, 57(2), s. 361–381. DOI: 10.1007/s10462-023-10641-x.

Suganya, V. Sasirekha, V. (2024). Economic Implications of Virtual Goods and Digital Assets in the Scope of Metaverse: An Analysis. W: *Omnichannel Approach to Co-Creating Customer Experiences Through Metaverse Platforms* (s. 32–38), B. Singla, K. Shalender, N. Singh (Eds.). Hershey: IGI Global. DOI: 10.4018/979-8-3693-1866-9.ch003O.

Szpringer, W. (2023). *Metawersum: nowe wyzwania dla zarządzania w gospodarce cyfrowej*. Warszawa: Wydawnictwo Poltext.

Upadhyay, A.K., Khandelwal, K. (2019). Artificial Intelligence-Based Training Learning from Application, *Development and Learning in Organizations: An International Journal*, 33(2), s. 20–23. DOI: 10.1108/dlo-05-2018-0058.

Wang, G., Badal, A., Jia, X., Maltz, J.S., Mueller, K., Myers, K.J., Niu, C., Vannier, M., Yan, P., Yu, Z., Zeng, R. (2022). Development of Metaverse for Intelligent Healthcare, *Nature Machine Intelligence*, 4(12), s. 1173–1180. DOI: 10.1038/s42256-022-00549-6.

Zhao, Y., Jiang, J., Chen, Y., Liu, R., Yang, Y., Xue, X., Chen, S. (2022). Metaverse: Perspectives from Graphics, Interactions and Visualization, *Visual Informatics*, 6(1), s. 56–67.

16.0

ZASTOSOWANIE SZTUCZNEJ INTELIGENCJI W DZIAŁANIACH ZMIERZAJĄCYCH DO OGRANICZENIA KONSUMPCJI TYTONIU – ANALIZA PRZYPADKÓW NHS I WHO

Marcin Hofman
Szkoła Doktorska przy Wydziale Ekonomicznym Uniwersytetu Gdańskiego
Jacek Winiarski
Uniwersytet Gdański

Streszczenie

Konsumpcja wyrobów tytoniowych prowadzi do znaczących obciążeń zarówno dla budżetu poszczególnych państw, jak i społeczeństwa jako całości. Innowacyjne technologie, takie jak sztuczna inteligencja, mogą wzbogacić programy antynikotynowe w organizacjach zdrowia publicznego, podnosząc ich efektywność. Rozdział opisuje sposoby, w jakie jedne z największych i najbardziej innowacyjnych organizacji zdrowia publicznego – WHO i NHS używają sztucznej inteligencji w swoich narzędziach. Przywołane organizacje, wykorzystując technologie AI, efektywnie zmniejszają koszty leczenia, zwiększają efektywność swoich działań oraz prowadzą skuteczne programy edukacyjne i prewencyjne. Działania te nie tylko poprawiają zdrowie publiczne, ale także znacząco redukują obciążenia finansowe związane z konsekwencjami palenia tytoniu. W niniejszym rozdziale podjęta została również próba przedstawienia wyliczeń w postaci kosztów i opcjonalnych oszczędności. Opisano także liczne wyzwania

oraz obawy społeczne, jakie niesie za sobą implementacja sztucznej inteligencji w systemach służby zdrowia.

Słowa kluczowe: sztuczna inteligencja, organizacje zdrowia publicznego, koszty społeczno-ekonomiczne, innowacyjność

Wprowadzenie

Według szacunków Światowej Organizacji Zdrowia (WHO) ponad miliard ludzi na całym świecie korzysta z wyrobów tytoniowych, m.in. papierosów, a ponad 80% z tych osób żyje w krajach o niskim lub średnim dochodzie. Wysokie koszty ekonomiczne, jakie ponoszą kraje w związku z nałogiem palenia, można pokazać na przykładzie Polski, gdzie w 2023 r. budżet Narodowego Funduszu Zdrowia wyniósł ok. 142 mld PLN, a przybliżony bezpośredni koszt związany z leczeniem palaczy – 50 mld PLN [PAP, 2023]. W związku z tak znamiennym obciążeniem ekonomicznym organizacje służby zdrowia, finansowane m.in. ze środków publicznych, wdrażają kompleksowe polityki w celu złagodzenia skutków ekonomicznych i zdrowotnych związanych z omawianym nałogiem. Narzędzia oferowane przez dynamicznie rozwijającą się sztuczną inteligencję mogą znacząco wspierać walkę z paleniem tytoniu. Dzięki analizie danych, personalizacji interwencji, optymalizacji kampanii społecznych, wspieraniu decyzji dotyczących polityki zdrowotnej, monitorowaniu skutków oraz interakcji z użytkownikami sztuczna inteligencja może pomóc w opracowywaniu skuteczniejszych strategii antynikotynowych. Celem niniejszego rozdziału jest określenie wpływu implementacji sztucznej inteligencji na systemy służby zdrowia, ze szczególnym uwzględnieniem działań takich organizacji, jak WHO i NHS. A zatem omówiono w nim teoretyczne wyzwania i korzyści ekonomiczne związane z tą technologią, zwracając uwagę na jej rolę w poprawie skuteczności programów zdrowotnych oraz strategii antynikotynowych.

16.1. Koszty społeczno-ekonomiczne palenia wyrobów tytoniowych

Palenie tytoniu stanowi ogromny problem zdrowotny i ekonomiczny, generując nie tylko wysokie koszty leczenia chorób związanych z paleniem, ale także powodując znaczne straty związane ze spadkiem wydajności pracy i przedwczesnymi zgonami, co wpływa negatywnie na gospodarki i struktury społeczne na całym świecie [Parry,

Small, 2002]. Faktem jest, że np. w Polsce dochody z podatków akcyzowych są jednym z ważniejszych źródeł dochodów budżetowych (ok. 25 mld PLN w 2022 r.), natomiast wszelkie koszty, takie jak: wydatki na leczenie, przedwczesna śmierć, absencja chorobowa, obciążenie systemu opieki społecznej oraz te związane z kontrolą i regulacją rynku tytoniowego sprawiają, że całkowity ciężar spowodowany ryzykiem utraty zdrowia, związanym z produktami tytoniowymi, przewyższa wszelkie korzyści ekonomiczne z ich produkcji oraz sprzedaży [Jha, Peto, 2014]. Prawidłowość ta jest szczególnie widoczna w krajach o wysokim dochodzie, takich jak Wielka Brytania czy Australia [Barnes *et al.*, 2022]. Natomiast w krajach o niskim dochodzie ograniczone zasoby do monitorowania i zarządzania zdrowiem mogą sprawić, że pełny ciężar kosztów jest mniej widoczny, ale wciąż istotny. W tabeli 16.1 przedstawiono koszty (z podziałem na rodzaje i ich składowe), jakie ponosi Wielka Brytania w związku z konsumpcją przez część jej mieszkańców wyrobów tytoniowych.

Tabela 16.1. Całkowite koszty wynikające z konsumpcji wyrobów tytoniowych w Wielkiej Brytanii w 2022 r.

Rodzaj kosztu	Koszt (GBP)
Koszty produktywności	**32,0 bln**
▪ Straty dochodów związane z paleniem	9,3 bln
▪ Bezrobocie związane z paleniem	7,3 bln
▪ Wczesne zgony związane z paleniem	1,8 bln
▪ Zmniejszona GVA z powodu wydatków na tytoń	13,6 bln
Koszty opieki zdrowotnej	**1,9 bln**
Koszty opieki społecznej	**15,0 bln**
▪ Koszt opieki domowej	644,3 mln
▪ Koszt opieki w placówkach	588,1 mln
▪ Koszt nieformalnej opieki przez rodzinę i przyjaciół	8,4 bln
▪ Koszt niezaspokojonych potrzeb opiekuńczych	5,4 bln
Koszty pożarów	328,1 mln
▪ Koszt śmierci z powodu pożarów związanych z paleniem	137,2 mln
▪ Koszt obrażeń z powodu pożarów związanych z paleniem	84,4 mln
▪ Koszt szkód majątkowych z powodu pożarów związanych z paleniem	98,2 mln
▪ Roczny koszt dla służb ratowniczych z powodu pożarów związanych z paleniem	8,3 mln
Suma całkowitych kosztów w Wielkiej Brytanii	**49,2 bln**

Źródło: opracowanie własne na podstawie ASH [2023].

W tym samym roku przychody z podatków w Wielkiej Brytanii od wyrobów tytoniowych wyniosły £12,8 mld. Naturalnie, przychody z podatku akcyzowego to główne źródło dochodów budżetowych z tytułu sprzedaży papierosów, ale nie jedyne. Oprócz podatków akcyzowych, które są bezpośrednim źródłem wpływów finansowych, istnieją również inne potencjalne źródła dochodów związane z przemysłem tytoniowym, takie jak opłaty licencyjne, grzywny za naruszenia przepisów oraz wpływy z regulacji i kontrolowania rynku tytoniowego. W świetle przykładów z Wielkiej Brytanii można jednak stwierdzić, że całkowite koszty ekonomiczne związane z używaniem wyrobów tytoniowych przewyższają dochody z podatków. W 2022 r. globalne koszty palenia tytoniu wyniosły około 1,4 bln USD, co stanowi około 2% rocznego światowego PKB. W odpowiedzi na te wyzwania ASH (Action on Smoking and Health) podkreśla potrzebę skuteczniejszej polityki, w tym strategii eliminacji palenia w przyszłych generacjach, mogącej znacznie zmniejszyć liczbę palaczy.

16.2. Rola sztucznej inteligencji w zdrowiu publicznym

Sztuczna inteligencja (AI) jest opisywana jako zaawansowane systemy komputerowe, wykorzystujące różnorodne techniki i algorytmy do wykonywania zadań wymagających inteligencji ludzkiej [Lee, Chen, 2022], dąży zaś do rozwijania systemów i narzędzi, które mogą wykonywać zadania wymagające inteligencji, podobnie jak w przypadku działań realizowanych przez ludzi. Rozwiązywanie problemów, rozumowanie oraz rozumienie języka to przykłady zadań, które sztuczna inteligencja ma na celu realizować. Warto jednak zauważyć, że pojęcie uczenia się w kontekście AI odnosi się do procesu, w którym systemy komputerowe dostosowują swoje algorytmy i modele na podstawie analizowanych danych, aby efektywniej rozwiązywać konkretne problemy. Uczenie się w tym przypadku nie jest celem samym w sobie, lecz narzędziem umożliwiającym poprawę zdolności systemu do realizacji zadań, takich jak analiza danych, klasyfikacja czy przewidywanie. W literaturze przedmiotu podkreśla się, że sztuczna inteligencja jest transformacyjną technologią, mającą potencjał zrewolucjonizować niemal każdy aspekt społeczeństwa [Tegmark, 2017]. Technologia ta to także systemy służące do rozwiązywania problemów wymagających funkcji poznawczych. Innowacyjność sztucznej inteligencji sprawia, że znajduje ona zastosowanie w niemal każdej branży. Poprawne zastosowanie narzędzi oferowanych przez AI może prowadzić do redukcji kosztów, zwiększenia wydajności oraz innowacyjności [Russell, 2019].

Technologie sztucznej inteligencji mogą być skutecznie stosowane w obszarze zdrowia publicznego. Od 2020 r. zauważalny jest dynamiczny wzrost ich integracji w strategiach zdrowia publicznego [Shah, Patel, Williams, 2023]. Możliwość analizy ogromnych zbiorów danych medycznych stanowi jedną z kluczowych funkcji algorytmów uczenia maszynowego. Dokładna analiza takich danych pomaga w identyfikacji i prognozie trendów zdrowotnych, co prowadzi do lepszego planowania precyzyjnych interwencji medycznych. Sztuczna inteligencja zmienia zdrowie publiczne poprzez monitorowanie wybuchów epidemii, personalizowanie działań zapobiegawczych oraz optymalizowanie alokacji zasobów. Chociaż jej rola w odpowiedzi na sytuacje kryzysowe może być ograniczona, technologie te znacząco wspierają inne aspekty zarządzania zdrowiem publicznym.

Konkretny przykład wspomagania zdrowia publicznego przez AI to jej zastosowanie w pulmonologii. Algorytmy poprawiają szybkość i dokładność diagnostyki, umożliwiając lekarzom skuteczne działanie np. w wykrywaniu zatorowości płucnej. Z kolei w gastroenterologii implementacja sztucznej inteligencji podczas kolonoskopii zmniejsza o 50% wskaźnik pominięcia gruczolaków przez lekarzy [Spadaccini *et al.*, 2023]. AI ma również potencjał wspierania tworzenia polityki antynikotynowej, oferując nowe możliwości w analizie danych i opracowywaniu strategii poprzez personalizację programów wsparcia oraz monitorowania postępów ludzi w walce z nałogiem w czasie rzeczywistym.

Z perspektywy ekonomicznej implementacja sztucznej inteligencji w systemie opieki zdrowotnej może mieć znaczący wpływ na budżet państwa zarówno w kontekście potencjalnych korzyści, jak i kosztów. Z jednej strony, AI ma potencjał do poprawy efektywności opieki zdrowotnej poprzez optymalizację procesów diagnostycznych, automatyzację administracji i przewidywanie potrzeb pacjentów, co może prowadzić do obniżenia kosztów leczenia i poprawy jakości usług zdrowotnych. Przykładem może być wspomniane zastosowanie algorytmów uczenia maszynowego w diagnostyce obrazowej, które mogą zwiększyć dokładność diagnoz i zmniejszyć liczbę błędów medycznych. Z drugiej strony, wdrażanie sztucznej inteligencji wiąże się z wysokimi kosztami, związanymi z dużymi projektami badawczo-rozwojowymi. Przykładem są inwestycje w rozwój i integrację zaawansowanych systemów AI, takich jak inteligentne systemy zarządzania danymi pacjentów czy algorytmy do analizy dużych zbiorów danych medycznych [Davenport, Kalakota, 2019]. Takie projekty mogą wymagać znacznych nakładów finansowych oraz czasu, a nie zawsze kończą się sukcesem. Istnieje ryzyko, że inwestycje w AI mogą nie przynieść oczekiwanych rezultatów, co może wpływać na stabilność finansową systemu opieki zdrowotnej. Natomiast dzięki oszczędnościom, które można osiągnąć poprzez wdrożenie efektywnych rozwiązań, możliwe

jest złagodzenie obciążenia budżetu państwa i skuteczniejsze alokowanie środków na opiekę zdrowotną. Jest to szczególnie istotne w obliczu starzejącego się społeczeństwa, gdzie optymalizacja wydatków na zdrowie staje się kluczowa dla zapewnienia długoterminowej stabilności finansowej systemu opieki zdrowotnej.

16.3. Sztuczna inteligencja i programy antytytoniowe. Studia przypadków – NHS i WHO

National Health Service (NHS) oraz Światową Organizację Zdrowia (WHO) można określić jako instytucje zdrowia publicznego, gdyż obie działają na rzecz poprawy zdrowia ludności. Pierwsza z nich, założona w 1948 r., działa na poziomie krajowym w Wielkiej Brytanii, natomiast WHO jest wyspecjalizowaną instytucją Organizacji Narodów Zjednoczonych, funkcjonującą na poziomie globalnym. Obydwie organizacje współpracują ze sobą m.in. w reagowaniu na kryzysy zdrowotne (np. pandemia COVID-19), w kampaniach zachęcających do szczepień czy też programach zdrowotnych w obszarze chorób przewlekłych. Współpraca w zakresie sztucznej inteligencji także łączy obie jednostki. Doskonale pokazuje to cytat z dokumentu NHS: „NHS współpracuje z międzynarodowymi organizacjami, takimi jak WHO, w celu wdrażania najnowszych technologii i sztucznej inteligencji w opiece zdrowotnej. Taka współpraca pozwala na testowanie i implementację innowacyjnych rozwiązań, które mają na celu poprawę jakości opieki oraz efektywności systemu zdrowotnego" [NHS England, 2022].

Ścisła współpraca pomiędzy WHO a NHS jest szczególnie znamienna także w kwestii niwelowania szkód zdrowotnych, społecznych i ekonomicznych, związanych z paleniem tytoniu. Globalne wytyczne opracowywane przez organizację działającą w skali globalnej, takie jak „ramowa konwencja o kontrolowaniu tytoniu", zostały dostosowane przez NHS do poziomu krajowego Wielkiej Brytanii. Międzynarodowe konferencje i seminaria, w których uczestniczą przedstawiciele zarówno WHO, jak i NHS odgrywają kluczową rolę w wymianie innowacyjnych rozwiązań oraz sprawdzonych praktyk w zakresie polityk antynikotynowych. Takie wydarzenia nie tylko umożliwiają transfer wiedzy na temat skutecznych strategii kontroli tytoniu, ale także sprzyjają dostosowywaniu i ulepszaniu polityk w odpowiedzi na nowe wyzwania i zmieniające się warunki zdrowotne. Współpraca międzynarodowa w tym zakresie prowadzi do opracowania bardziej zaawansowanych i zintegrowanych podejść do zwalczania palenia, co może znacząco wpłynąć na efektywność globalnych wysiłków w obszarze zdrowia publicznego [WHO, NHS. 2023]. Dalsza nieustanna współpraca i zrozumienie potrzeby innowacji w dostosowywaniu strategii antynikotynowych i personaliza-

cji programów dla osób uzależnionych są kluczowe dla skutecznej redukcji palenia tytoniu i ochrony zdrowia publicznego. Tradycyjne, stosowane od lat metody kontroli konsumpcji tytoniu obejmują: podwyższanie podatków na wyroby tytoniowe, zakazy palenia w miejscach publicznych, wymóg umieszczania wyraźnych ostrzeżeń zdrowotnych na opakowaniach, a także regulacje dotyczące reklamy wyrobów tytoniowych. Faktem jest, że tego typu działania leżą w gestii rządów poszczególnych krajów, jednak organizacje zdrowia publicznego wpływają na te decyzje, promując działania w zakresie polityk antynikotynowych. Programy wsparcia dla osób rzucających konsumpcję wyrobów tytoniowych czy kampanie społeczne leżą bardziej w gestii służby zdrowia, takich jak NHS.

W dynamicznie zmieniającym się XXI w. organizacje zdrowia publicznego muszą mierzyć się z wieloma wyzwaniami, które utrudniają skuteczność kampanii i programów antynikotynowych. Jednym z największych wyzwań są agresywne kampanie promocyjne ze strony koncernów tytoniowych w kwestii promocji np. papierosów elektronicznych o różnorodnych smakach, użytkowanych przez coraz większą grupę ludzi w wieku 20–30 lat [Truth Initiative, 2023]. W odpowiedzi na te wyzwania WHO i NHS integrują innowacyjne rozwiązania, takie jak sztuczna inteligencja, w swoich programach zdrowotnych [NHS England, 2022]. Dzięki zastosowaniu narzędzi AI programy te mogą zyskać dodatkowe korzyści w postaci większej efektywności, lepszej personalizacji i szybszego dostosowywania się do potrzeb użytkowników, co może zwiększyć ich dostępność, skalowalność i użyteczność. Skutkować to może zmniejszeniem społeczno-ekonomicznych kosztów palenia oraz poprawą zdrowia publicznego na poziomie zarówno krajowym, jak i globalnym.

Jednym z najważniejszych działań Światowej Organizacji Zdrowia z zastosowaniem sztucznej inteligencji w niwelowaniu szkód związanych z konsumpcją wyrobów tytoniowych było wprowadzenie w 2020 r. wirtualnej asystentki zdrowotnej o imieniu Florence (rysunek 16.1).

To innowacyjne narzędzie, oparte na sztucznej inteligencji, jest częścią inicjatywy WHO wykorzystującej nowoczesne technologie do poprawy zdrowia publicznego na całym świecie. Cechy i funkcje Florence opierają się na algorytmach sztucznej inteligencji, a także na integracji z różnymi platformami i aplikacjami zdrowotnymi. Jedną z najważniejszych funkcji wirtualnej asystentki jest możliwość personalizacji planów rzucenia palenia wyrobów tytoniowych. Umiejętność analizy danych dotyczących nawyków palenia pozwala użytkownikom na otrzymywanie zindywidualizowanych porad np. na temat metod radzenia sobie z głodem nikotynowym, a także dostosowywania planów do stopnia zaawansowania nałogu. Kluczowe jest, aby programy i narzędzia zdrowotne nie tylko umożliwiały użytkownikom edukację na temat obalania mitów,

takich jak błędne przekonanie o braku szkodliwości papierosów elektronicznych, ale również oferowały wsparcie emocjonalne tam, gdzie to możliwe i odpowiednie. Florence oferuje takie funkcje, m.in. dostarczając codziennych motywacyjnych wiadomości, co ma zwiększyć determinację w kontynuacji procesu rzucenia palenia.

Rysunek 16.1. Florence, wirtualna asystentka wprowadzona przez WHO

Źródło: WHO [2020].

Jednym z istotnych problemów służby zdrowia jest długi czas oczekiwania na wizytę u lekarza-specjalisty, a także ograniczony czas, jaki medyk jest w stanie poświęcić pacjentowi, co może prowadzić do całkowitej rezygnacji z porady lekarskiej. Kluczową rolą narzędzi opartych na sztucznej inteligencji powinny więc być dostępność i interaktywność. Wirtualna asystentka, wprowadzona przez WHO, dzięki dostępności przez całą dobę oraz umiejętności prowadzenia z użytkownikami rozmowy, w sposób przypominający interakcje międzyludzkie, spełnia te kryteria. Światowa Organizacja Zdrowia w dalszym ciągu inwestuje w to narzędzie, rozszerzając bazę wiedzy, dodając nowe języki, co jest kluczowym elementem w przeciwdziałaniu epidemii palenia tytoniu w skali globalnej.

Organizacja zdrowia publicznego w Wielkiej Brytanii, podobnie jak WHO, stosuje sztuczną inteligencję w swoich programach antynikotynowych, jednak na poziomie krajowym. Jednym z najbardziej zaawansowanych narzędzi w tej dziedzinie jest cyfrowa platforma zdrowotna – Quit Genius. Narzędzie to oferuje zaawansowane funkcje, które na podstawie indywidualnych nawyków użytkowników opracowują spersonali-

zowane plany leczenia. Program wyróżnia się także różnorodnymi ćwiczeniami behawioralnymi, które są technikami mającymi na celu modyfikację nawyków i zachowań użytkowników. Ćwiczenia te mogą obejmować strategie radzenia sobie z pokusami, rozwijanie zdrowszych nawyków oraz naukę technik relaksacyjnych i radzenia sobie ze stresem, wspierające proces rzucania palenia. Zasadniczo aplikacja pod względem funkcjonalności jest podobna do wirtualnej asystentki zdrowotnej WHO, gdyż także pomaga utrzymać wysoki poziom zaangażowania czy też monitorować postępy w czasie rzeczywistym. Aby określić, w jakim stopniu narzędzie to przynosi zaplanowane rezultaty, przeprowadzono odpowiednie badania, po analizie których przedstawiono następujące fakty [Quit Genius, 2022]:

- szansa na rzucenie palenia przy użyciu aplikacji w porównaniu z tradycyjnymi metodami zwiększa się o 52%,
- korzystanie z aplikacji przez co najmniej miesiąc, co świadczy o wysokim poziomie zaangażowania, kontynuuje 75% użytkowników,
- zmniejszenie objawu głodu nikotynowego odczuwa 65% użytkowników.

Wyniki te sugerują, że Quit Genius jest skutecznym narzędziem wspomagającym rzucenie palenia, choć do pełnego porównania z innymi metodami potrzebne są dalsze badania.

Kolejnym zaawansowanym narzędziem AI, wspieranym przez NHS, jest aplikacja mobilna Smoke Free, której interfejs na urządzeniach mobilnych został zaprezentowany na rysunku 16.2.

Aplikacja jest dostępna na urządzenia mobilne z systemem iOS i Android. Umożliwia łączenie się z innymi osobami, które również dążą do rzucenia palenia, co wspiera tworzenie relacji międzyludzkich. Interfejs aplikacji jest prosty w obsłudze i łatwy do zrozumienia, a regularne aktualizacje pomagają utrzymać jej funkcjonalność na wysokim poziomie.

Przeprowadzone badania kontrolne i uzyskane wyniki [Crane *et al.*, 2018] wskazują, że użytkownicy aplikacji Quit Genius mają o 90% większe szanse na rzucenie palenia niż osoby stosujące tradycyjne metody, a wynik ten jest statystycznie istotny ($p < 0{,}001$) i potwierdzony wysokim poziomem pewności (95% CI = 1,53–2,37). Oznacza to, że aplikacja znacząco zwiększa skuteczność w walce z nałogiem. Niezależni eksperci przyznali omawianemu oprogramowaniu 91% zgodności z najlepszymi praktykami medycznymi, co czyni ją najczęściej rekomendowaną aplikacją wśród brytyjskich lekarzy.

Podobną aplikacją antytytoniową tego typu jest EX Program, wykorzystujący sztuczną inteligencję do personalizacji wsparcia dla użytkowników. Główną innowację stanowi algorytm rekomendacyjny, który analizuje motywacje i preferencje

użytkownika, dostarczając mu treści dopasowane do jego indywidualnych potrzeb. Aplikacja łączy się z autorytetami w zakresie uzależnień od nikotyny, takimi jak Mayo Clinic, aby dostarczać skuteczne, oparte na badaniach strategie rzucania palenia.

Rysunek 16.2. Aplikacja Smoke Free

Źródło: Smoke Free (2024).

Aplikacje antytytoniowe zintegrowane z AI wnoszą istotną wartość dla użytkowników, oferując zaawansowane funkcje, które wyróżniają je na tle aplikacji bez takiej technologii. Przede wszystkim AI pozwala na personalizację i dopasowanie wsparcia do indywidualnych potrzeb użytkowników, co znacząco poprawia skuteczność takich aplikacji. Tradycyjne aplikacje często dostarczają ogólne informacje i porady, natomiast aplikacje z AI mogą reagować dynamicznie, dostosowując wsparcie do konkretnej sytuacji i wcześniejszych zachowań użytkownika. Algorytmy w aplikacjach opartych na AI analizują preferencje użytkownika i jego motywacje, aby dostarczać spersonalizowane treści, takie jak artykuły, porady czy plany rzucania palenia, dostosowane do indywidualnych potrzeb. Jest to rozwiązanie, które zwiększa zaangażowanie użytkowników, ponieważ treści są bardziej adekwatne do ich sytuacji i poziomu zaawansowania w procesie rzucania.

Podsumowując, aplikacje i programy, stosowane jako pomoc w zaprzestaniu konsumpcji wyrobów tytoniowych, wykorzystują innowacyjne rozwiązania oparte na sztucznej inteligencji i mają potencjał do znacznego zmniejszenia szkód społeczno-ekonomicznych.

16.4. Wyzwania związane z wykorzystaniem sztucznej inteligencji w organizacjach zdrowia publicznego

Obok ogromnych korzyści związanych z wdrażaniem sztucznej inteligencji w organizacjach zdrowia publicznego wiążą się z nią również liczne wyzwania [Lee,Chen, 2022]. Jednym z największych jest skuteczna ochrona prywatności danych pacjentów [Russell, 2019]. W literaturze przedmiotu zwraca się uwagę na konieczność zapewnienia transparentności algorytmów oraz skutecznego zarządzania złożonością integracji systemów AI z istniejącą infrastrukturą opieki zdrowotnej [Bohr, Memarzadeh, 2020]. Ponadto inny autor zwraca uwagę na etyczne rozważania, takie jak ryzyko uprzedzeń w algorytmach [Tegmark, 2017]. Dodatkowym wyzwaniem są zmieniające się regulacje prawne [Reddy, 2020]. Narzędzia i aplikacje stosujące rozwiązania AI muszą być zgodne nie tylko z międzynarodowymi, ale także lokalnymi regulacjami dotyczącymi zdrowia, aby mogły być one stosowane w skali międzynarodowej [Shah, Patel, Williams, 2023]. Kosztownymi i czasochłonnymi wyzwaniami są także szkolenia personelu medycznego oraz rygorystyczne procesy certyfikacji i akredytacji.

Podsumowanie

W rozdziale przeanalizowano zastosowanie sztucznej inteligencji (AI) w programach zdrowia publicznego, mających na celu redukcję konsumpcji wyrobów tytoniowych. Najważniejsze wnioski wskazują na wysoką skuteczność narzędzi AI, takich jak wirtualna asystentka Florence (WHO) oraz aplikacja Quit Genius (NHS), które zwiększają zaangażowanie użytkowników oraz ich szanse na trwałe zaprzestanie palenia tytoniu. Personalizacja interwencji oraz stały dostęp do wsparcia stanowią kluczowe czynniki sukcesu tych narzędzi. Korzyści obejmują również redukcję kosztów leczenia chorób związanych z paleniem.

W rozdziale zwrócono także uwagę na wyzwania związane z wdrażaniem AI, w tym ochronę danych pacjentów, wysokie koszty wdrożenia oraz kwestie etyczne dotyczące

transparentności algorytmów. Dynamicznie zmieniające się regulacje prawne stanowią dodatkowy czynnik komplikujący proces implementacji.

Ograniczenia badań obejmują potrzebę bardziej szczegółowych analiz długoterminowych skutków stosowania AI w programach zdrowia publicznego oraz konieczność monitorowania skuteczności w różnych kontekstach kulturowych i społecznych. W przyszłych badaniach warto zwrócić uwagę na szersze zastosowanie AI w innych obszarach zdrowia publicznego oraz dokładniejsze oszacowanie kosztów i korzyści ekonomicznych wynikających z jej wdrożenia.

Bibliografia

ASH (2023). *New Figures Show Smoking Costs Billions More than Tobacco Taxes as Consultation on Creating a Smokefree Generation Closes*, https://ash.org.uk/media-centre/news/press-releases/new-figures-show-smoking-costs-billions-more-than-tobacco-taxes-as-consultation-on-creating-a-smokefree-generation-closes (dostęp: 27.07.2024).

Barnes, R.L., Yach, D., Jha, P. (2022). Global Economic Cost of Smoking-Attributable Diseases, *Tobacco Control*, 31(1), s. 1–12.

Bohr, A., Memarzadeh, K. (2020). *Artificial Intelligence in Healthcare*. Cambridge: Academic Press.

Crane, D., Ubhi, H.K., Brown, J., West, R. (2018). Relative Effectiveness of a Full Versus Reduced version of the 'Smoke Free' Mobile Application for Smoking Cessation: An Exploratory Randomised Controlled Trial, *F1000Research*, *7*, 1524. DOI: 10.12688/f1000research.16148.2.

Davenport, T.H., Kalakota, R. (2019). The Potential for Artificial Intelligence in Healthcare, *Future Healthcare Journal*, 6(2), s. 94–98. DOI: 10.7861/futurehosp.6-2-94.

Jha, P., Peto, R. (2014). Global Burden of Disease Attributable to Tobacco and Alcohol, *The Lancet*, 384(9940), s. 987–998. DOI: 10.1016/S0140-6736(14)60696-0.

Lee, K.-F., Chen, Q. (2022). *AI 2041: Ten Visions for Our Future*. New York: Currency.

McDaid, D., Sassi, F., Merkur, S. (2017). Return on Investment of Public Health Interventions: a Systematic Review, *Journal of Epidemiology and Community Health*, 71(8), s. 827–834.

Mitchell, M. (2019). *Artificial Intelligence: A Guide for Thinking Humans*. New York: Farrar, Straus and Giroux.

NHS England (2022). *NHS Long Term Plan: Digital Transformation*, https://www.england.nhs.uk/long-term-plan/ (dostęp:27.07.2024).

PAP (2023). *Papierosy dziurawią polski budżet na 9,2 mld zł rocznie*, https://www.pap.pl/aktualnosci/news%2C1600923%2Cpapierosy-dziurawia-polski-budzet-na-92-mld-zl-rocznie.html (dostęp:27.07.2024).

Parry, I.W.H., Small, K.A. (2002). *Economics of Tobacco Control: Towards an Optimal Policy Mix*. Washington, D.C.: World Bank.

Quit Genius (2022). *New Randomized-Controlled Trial Finds That Quit Genius Digital Clinic Is Significantly More Effective Than Traditional Standard of Care for Addiction*, https://www.globenewswire.com/en/news-release/2022/05/05/2436926/0/en/New-Randomized-Con-

trolled-Trial-Finds-that-Quit-Genius-Digital-Clinic-is-Significantly-More-Effective-than-Traditional-Standard-of-Care-for-Addiction.html (dostęp: 27.07.2024).

Reddy, S. (2020). *Artificial Intelligence and Machine Learning in Public Health*. Singapore: Springer.

Russell, S. (2019). *Human Compatible: Artificial Intelligence and the Problem of Control*. New York: Viking.

Shah, P., Patel, V., Williams, J. (Eds.). (2023). *Artificial Intelligence in Healthcare*. Cambridge Academic Press.

Smith, J. (2024). How AI Can Revolutionize the Fight Against Smoking, *Journal of Digital Health Innovations*, 15(3), s. 45–59.

Smoke Free (2024). [Home], https://app.smokefreeapp.com/img/mobile-screen.png (dostęp: 30.07.2024).

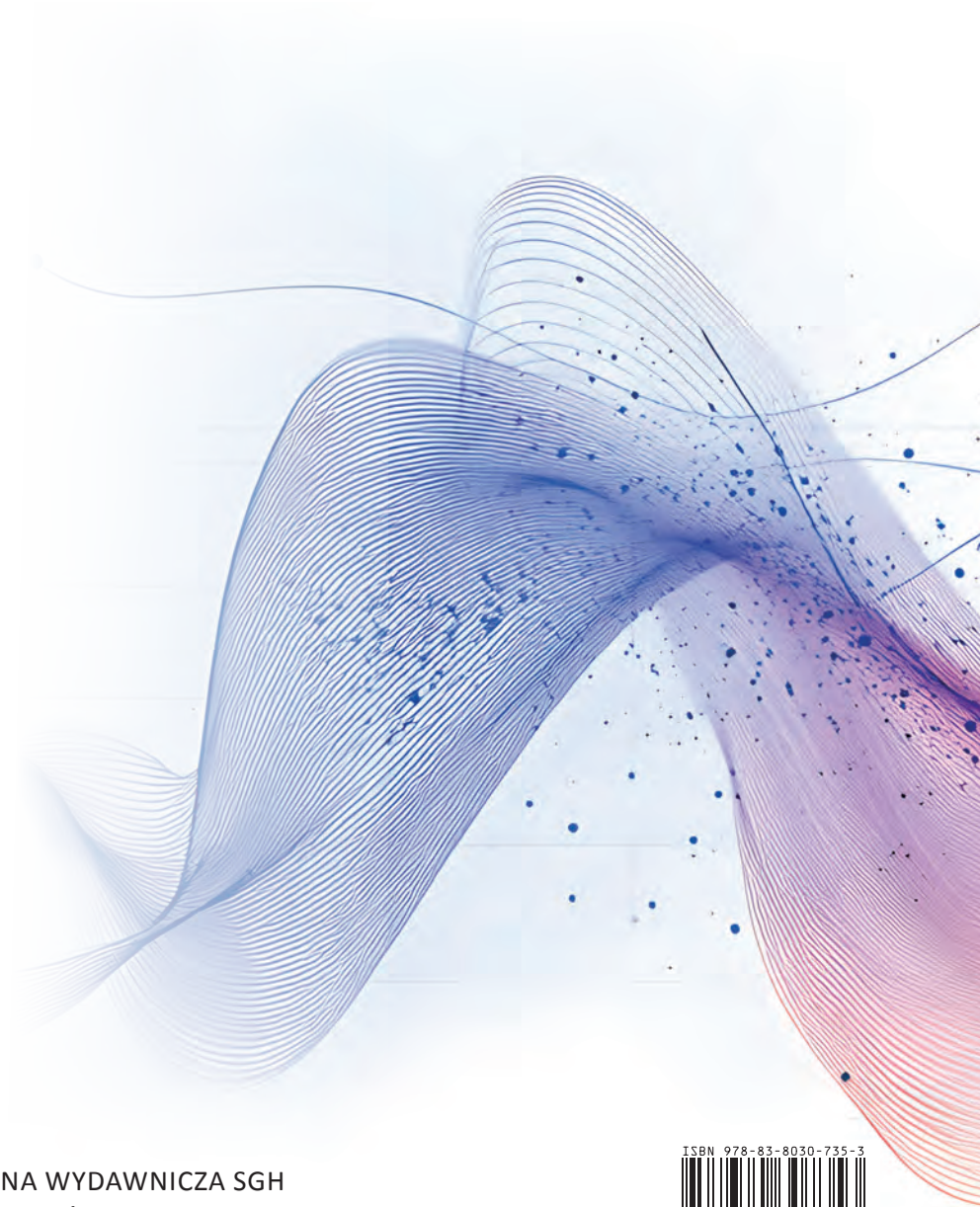
Spadaccini, M., Massimi, D., Mori, Y., Alfarone, L., Fugazza, A., Maselli, R., Sharma, P., Facciorusso, A., Hassan, C., Repici, A. (2023). Artificial Intelligence-Aided Endoscopy and Colorectal Cancer Screening, *Diagnostics (Basel)*, 13(6), 1102. DOI: 10.3390/diagnostics13061102.

Tegmark, M. (2017). *Life 3.0: Being Human in the Age of Artificial Intelligence*. New York: Alfred A. Knopf.

Truth Initiative. (2023). 2023 *Monitoring the Future Survey Shows Encouraging Declines in Youth E-Cigarette Use and Increased Risk Perception Among High Schoolers*, https://truthinitiative.org/press/press-release/2023-monitoring-future-survey-shows-encouraging-declines-youth-e-cigarette-use (dostęp: 28.07.2024).

WHO (2020). *Initiative Pour un Monde sans Tabac*, https://www.emro.who.int/fr/tfi/news/who-launches-arabic-florence-a-digital-health-worker-to-help-more-people-quit-tobacco.html (dostęp: 28.07.2024).

WHO, NHS (2023). *Global and Local Efforts in Tobacco Control*, https://www.who.int/publications/i/item/9789241565561 (dostęp: 28.07.2024).



OFICYNA WYDAWNICZA SGH
SZKOŁA GŁÓWNA HANDLOWA W WARSZAWIE
www.wydawnictwo.sgh.waw.pl

ISBN 978-83-8030-735-3