RODO dla AI

Zgodność z zasadami godnej zaufania sztucznej inteligencji w modelu *data protection by design*

Dominik Lubasz

Wolters Kluwer

RODO dla AI

Zgodność z zasadami godnej zaufania sztucznej inteligencji w modelu *data protection by design*

Dominik Lubasz

Wolters Kluwer

WARSZAWA 2025

SERIA **MONOGRAFIE**

Stan prawny na 20 stycznia 2025 r.

Recenzenci
Dr hab. Agnieszka Grzelak, prof. Akademii Leona Koźmińskiego
Dr hab. Mariusz Jagielski, prof. Uniwersytetu Śląskiego

Wydawczyni
Monika Pawłowska

Redaktorka prowadząca
Joanna Cybulska

Opracowanie redakcyjne
Katarzyna Świerk-Bożek

Projekt okładek serii
Wojtek Janikowski, Przemek Dębowski

Badania prowadzące do powstania tej monografii zostały częściowo sfinansowane przez Narodowe Centrum Nauki w ramach projektu badawczego „Ochrona konsumenta a sztuczna inteligencja. Między prawem a etyką" (nr projektu: 2018/31/B/HS5/01169), wybranego w konkursie OPUS 16, którego jednym z wykonawców jest autor.

ISBN 978-83-8390-361-3 *ISBN PDF-a: 978-83-8390-362-0*
ISSN 1897-4392

Wolters Kluwer Polska Sp. z o.o.
Dział Praw Autorskich
01-208 Warszawa, ul. Przyokopowa 33
tel. +48 728 313 462
e-mail: PL-ksiazki@wolterskluwer.com

księgarnia internetowa www.profinfo.pl

SPIS TREŚCI

WYKAZ SKRÓTÓW

Akty prawne

AIA	–	rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z 13.06.2024 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) (Dz.Urz. UE L 2024/1689)
DMA	–	rozporządzenie Parlamentu Europejskiego i Rady (UE) 2022/1925 z 14.09.2022 r. w sprawie kontestowalnych i uczciwych rynków w sektorze cyfrowym oraz zmiany dyrektyw (UE) 2019/1937 i (UE) 2020/1828 (akt o rynkach cyfrowych) (Dz.Urz. UE L 265, s. 1)
DSA	–	rozporządzenie Parlamentu Europejskiego i Rady (UE) 2022/2065 z 19.10.2022 r. w sprawie jednolitego rynku usług cyfrowych oraz zmiany dyrektywy 2000/31/WE (akt o usługach cyfrowych) (Dz.Urz. UE L 277, s. 1 ze zm.)
EKPC	–	Konwencja o ochronie praw człowieka i podstawowych wolności sporządzona w Rzymie 4.11.1950 r. (Dz.U. z 1993 r. Nr 61, poz. 284 ze zm.)
Konstytucja RP	–	Konstytucja Rzeczypospolitej Polskiej z 2.04.1997 r. (Dz.U. Nr 78, poz. 483 ze zm.)
KPP	–	Karta praw podstawowych Unii Europejskiej (Dz.Urz. UE C 202 z 2016 r., s. 389)

RODO	–	rozporządzenie Parlamentu Europejskiego i Rady (UE) 2016/679 z 27.04.2016 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) (Dz.Urz. UE L 119, s. 1 ze sprost.)
rozporządzenie 1019/1020	–	rozporządzenie Parlamentu Europejskiego i Rady (UE) 2019/1020 z 20.06.2019 r. w sprawie nadzoru rynku i zgodności produktów oraz zmieniające dyrektywę 2004/42/WE oraz rozporządzenia (WE) nr 765/2008 i (UE) nr 305/2011 (Dz.Urz. UE L 169, s. 1 ze zm.)
TFUE	–	Traktat o funkcjonowaniu Unii Europejskiej (Dz.Urz. UE C 202 z 2016 r., s. 47)
TUE	–	Traktat o Unii Europejskiej (Dz.Urz. UE C 202 z 2016 r., s. 13)

Czasopisma

EPS	–	Europejski Przegląd Sądowy
M. Praw.	–	Monitor Prawniczy

Inne

AI	–	*Artificial Intelligence* (sztuczna inteligencja)
CEN	–	Comité européen de normalisation (Europejski Komitet Normalizacyjny)
CENELEC	–	Comité européen de normalisation electrotechnique (Europejski Komitet Normalizacyjny Elektrotechniki)
CNIL	–	Commission Nationale de l'Informatique et des Libertés (francuska Krajowa Komisja ds. Technologii Informacyjnych i Wolności)
ENISA	–	European Union Agency for Cybersecurity (Agencja Unii Europejskiej ds. Cyberbezpieczeństwa)
EROD	–	Europejska Rada Ochrony Danych
ETPC	–	Europejski Trybunał Praw Człowieka
HLEG AI	–	High-Level Expert Group on Artificial Intelligence (Grupa ekspertów wysokiego szczebla ds. sztucznej inteligencji)
IEEE	–	Institute of Electrical and Electronics Engineers (Instytut Inżynierów Elektryków i Elektroników)

ISO	–	International Standards Organization (Międzynarodowa Organizacja Normalizacyjna)
NIST	–	National Institute of Standards and Technology (amerykański Narodowy Instytut Norm i Technologii)
NSA	–	Naczelny Sąd Administracyjny
OECD	–	Organisation for Economic Co-operation and Development (Organizacja Współpracy Gospodarczej i Rozwoju)
ONZ	–	Organizacja Narodów Zjednoczonych
TS/TSUE	–	Trybunał Sprawiedliwości Unii Europejskiej (przed 1.12.2009 r. jako Trybunał Sprawiedliwości)
UNESCO	–	Organizacja Narodów Zjednoczonych do spraw Edukacji, Nauki i Kultury
UODO	–	Urząd Ochrony Konkurencji i Konsumentów
WSA	–	Wojewódzki Sąd Administracyjny

WPROWADZENIE

1. Kontekst badawczy

Od wizji łoża Prokrusta[1] po dzień zagłady[2] – sztuczna inteligencja, jako siła transformująca rzeczywistość, stała się katalizatorem najgłębszych debat etycznych i egzystencjalnych, zmuszając do zdefiniowania zasad i wartości, które mają kierować jej dalszym rozwojem i wpływem tak na człowieka, jak i na całe społeczeństwa[3].

---

[1] Negatywna postać z mitologii greckiej znana pod tym przydomkiem, o imieniu Damastes, do której w literaturze odwoływali się również m.in. Z. Herbert w wierszu *Damastes z przydomkiem Prokrustes mówi* czy S. Lem w powieści *Eden* (S. Lem, *Eden*, Warszawa 2012) w kontekście zjawisk przymusowego dopasowywania jednostek lub całego społeczeństwa do określonych wzorców.

[2] Tezy takie formułował m.in. Stephen Hawking – zob. *Stephen Hawking warns artificial intelligence could end mankind*, https://www.bbc.com/news/technology-30290540 (dostęp: 16.12.2024 r.); zob. też K. Roose, *A.I. Poses 'Risk of Extinction', Industry Leaders Warn*, „The New York Times" 30.05.2023, https://www.nytimes.com/2023/05/30/technology/ai-threat-warning.html (dostęp: 16.12.2024 r.); W. Hartzog, *Two AI Truths and a Lie*, „Yale Journal of Law & Technology" 2024/26(3), s. 612.

[3] A. Jobin, M. Ienca, E. Vayena, *Artificial Intelligence: the global landscape of ethics guidelines*, https://www.researchgate.net/publication/334082218_Artificial_Intelligence_the_global_landscape_of_ethics_guidelines (dostęp: 16.12.2024 r.); E. Awad i in., *The Moral Machine experiment*, „Nature" 2018/563, https://www.researchgate.net/publication/341827530_2018_Article_The_Moral_Machine (dostęp: 16.12.2024 r.), s. 59–63; T. Timan, Z. Mann, *Data Protection in the Era of Artificial Intelligence. Trends, Existing Solutions and Recommendations for Privacy-preserving Technologies* [w:] *The Elements of Big Data Value, Foundations of the Research and Innovation Ecosystem*, red. E. Curry, A. Metzger, S. Zillner, J.-C. Pazzaglia, A. García Robles, Cham 2021, s. 165 i n.; J. Fjeld, N. Achten, H. Hilligoss, A.C. Nagy, M. Srikumar, *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3518482 (dostęp: 16.12.2024 r.).

W dyskusję tę zaangażowali się interesariusze z wielu środowisk, naukowcy z różnych dziedzin, działacze społeczni, przedsiębiorcy i ich reprezentanci, wreszcie państwa i organizacje międzynarodowe[4]. Jej głównym celem było zdefiniowanie wyzwań i dylematów natury etycznej, związanych z rozwojem i aplikacją systemów AI, a następnie adresujących te wyzwania zasad, wytycznych i regulacji. Do dyskusji tej włączyła się również Unia Europejska i jej państwa członkowskie w ramach działań wewnętrznych, jak i gremiów międzynarodowych, w szczególności OECD i Rady Europy.

Na płaszczyźnie unijnej w 2018 r. Komisja Europejska przedstawiła strategię AI, formułując koncepcję humanocentrycznej AI (*human-centric AI*), zgodnie z którą w Unii Europejskiej mają być poszukiwane, a następnie zdefiniowane i przyjęte rozwiązania regulacyjne, gwarantujące, że człowiek będzie stawiany w centrum rozwoju AI[5]. Pierwszym krokiem do osiągnięcia tego celu było powołanie Grupy ekspertów wysokiego szczebla ds. sztucznej inteligencji (High-Level Expert Group on Artificial Intelligence, HLEG AI), której powierzono zadanie opracowania podstawowych zasad etycznych, a następnie wytycznych dotyczących humanocentrycznej AI. W opracowanych wytycznych HLEG AI wskazała na konieczność stworzenia w Unii Europejskiej warunków do rozwoju AI skoncentrowanej na europejskich wartościach humanistycznych, przede wszystkim przez nadanie jej przymiotu „godna zaufania" (*trustworthy*). Godna zaufania AI powinna posiadać określone właściwości, które koncentrują się wokół zapewnienia gwarancji autonomii i kontroli, a także ochrony człowieka poddanego wpływowi zautomatyzowanych procesów decyzyjnych wykorzystujących AI[6]. Sformułowane przez HLEG AI zasady etyczne i związane z nimi wymogi dotyczące godnej zaufania

[4] Dyskusje te i ich efekty zostały omówione w rozdziale I pkt 2.3.

[5] Komunikat Komisji do Parlamentu Europejskiego, Rady Europejskiej, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów, Sztuczna inteligencja dla Europy, 25.04.2018 r., COM(2018) 237 final, https://eur-lex.europa.eu/legal-content/PL/TXT/PDF/?uri=CELEX:52018DC0237 (dostęp: 6.11.2023 r.).

[6] Szerzej zob. rozdział I pkt 3.2. Zob. też D. Lubasz, *Relacja ogólna aktu w sprawie sztucznej inteligencji i RODO* [w:] *Prawo sztucznej inteligencji i nowych technologii 3*, red. B. Fischer, A. Pązik, M. Świerczyński, Warszawa 2024, s. 157.

AI stały się, jako fundament deontologiczny, podstawą ukształtowania ram prawnych dla AI w Unii Europejskiej[7].

Analiza potrzeb regulacyjnych doprowadziła do przedstawienia przez Komisję Europejską w kwietniu 2021 r. projektu rozporządzenia ustanawiającego zharmonizowane zasady dotyczące sztucznej inteligencji (*Artificial Intelligence Act*, AIA)[8], które zostało przyjęte 13.06.2024 r. w zwykłej procedurze legislacyjnej i opublikowane 12.07.2024 r.[9] Akt ten, który miał ustanowić horyzontalne, zharmonizowane przepisy odnoszące się do AI, przewiduje jednak zasady mające zastosowanie do projektowania, rozwoju i wykorzystywania tylko niektórych systemów AI[10]. Powoduje to, że poza zakresem jego zastosowania jest znaczna część systemów AI i sposobów ich wykorzystywania, a ich normowanie odbywa się zasadniczo na podstawie dotychczas obowiązujących przepisów prawa, w tym RODO, przepisów konsumenckich, przepisów regulujących rynek cyfrowy, w tym DSA i DMA, prawa autorskiego, przepisów dotyczących bezpieczeństwa produktów, przepisów antydyskryminacyjnych i sektorowych, np. bankowych czy ubezpieczeniowych.

---

[7] Komunikat Komisji do Parlamentu Europejskiego, Rady, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów, Budowanie zaufania do sztucznej inteligencji ukierunkowanej na człowieka, 8.04.2019 r., COM(2019) 168 final, https://eur-lex.europa.eu/legal-content/PL/TXT/PDF/?uri=CELEX:52019DC0168 (dostęp: 17.12.2024 r.).

[8] Wniosek dotyczący rozporządzenia Parlamentu Europejskiego i Rady ustanawiającego zharmonizowane przepisy dotyczące sztucznej inteligencji (Akt w sprawie sztucznej inteligencji) i zmieniającego niektóre akty ustawodawcze Unii, 21.04.2021, COM(2021) 206 final, https://ec.europa.eu/transparency/documents-register/detail?ref=COM(2021)206&lang=pl (dostęp: 6.11.2023 r.).

[9] Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z 13.06.2024 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).

[10] Znalazło to również swój wyraz w motywie 9 zdanie drugie AIA, zgodnie z którym zharmonizowane przepisy ustanowione w AIA powinny mieć zastosowanie we wszystkich sektorach i – zgodnie z nowymi ramami prawnymi – powinny pozostawać bez uszczerbku dla obowiązującego prawa Unii, w szczególności w zakresie ochrony danych, ochrony konsumentów, praw podstawowych, zatrudnienia i ochrony pracowników oraz bezpieczeństwa produktów, wobec którego to prawa AIA ma charakter uzupełniający.

2. Uzasadnienie wyboru obszaru badawczego

Konsekwencją wąskiego zakresu zastosowania AIA jest konieczność poszukiwania w innych aktach prawnych instrumentów prawnych, które zapewnią realizację założeń ram prawnych dla humanocentrycznej AI w Unii Europejskiej. Poza deklaracją Komisji, że na szczeblu unijnym i krajowym istnieją już solidne ramy prawne mające na celu ochronę praw podstawowych oraz zapewnienie bezpieczeństwa i praw konsumentów, zwłaszcza RODO[11], a przyjęte podejście ma charakter proporcjonalny i oparte jest na analizie ryzyka, brak jest analizy, w jaki sposób wskazane dotychczasowe przepisy prawa mają się przyczynić do realizacji sformułowanych celów godnej zaufania AI.

Uwzględniając specyfikę tworzenia i funkcjonowania systemów AI, zwłaszcza w związku z zapotrzebowaniem na dane, spośród wskazanych regulacji to RODO wydaje się oferować największy potencjał oddziaływania w celu zapewnienia realizacji zasad godnej zaufania humanocentrycznej AI. Jest to związane z horyzontalnym zakresem zastosowania tego aktu oraz ogniskowaniem ochrony na prawach i wolnościach podmiotów danych z wykorzystaniem neutralnego technologicznie podejścia opartego na ryzyku, zasad przetwarzania i instrumentów uwzględniających ocenę z perspektywy adekwatnych regulacyjnie wartości (*values by design*)[12], przede wszystkim ochrony danych w fazie

[11] Komunikat Komisji do Parlamentu Europejskiego, Rady Europejskiej, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów, Sztuczna inteligencja dla Europy, 25.04.2018 r., COM(2018) 237 final, s. 16; Komunikat Komisji do Parlamentu Europejskiego, Rady Europejskiej, Rady, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów, Promowanie europejskiego podejścia do sztucznej inteligencji, 21.04.2021 r., COM(2021) 205 final, https://eur-lex.europa.eu/resource.html?uri=cellar:01ff45fa-a375-11eb-9585-01aa75ed71a1.0005.02/DOC_1&format=PDF (dostęp: 17.12.2024 r.), s. 7.

[12] Grupa ekspertów wysokiego szczebla ds. sztucznej inteligencji (HLEG AI), Wytyczne w zakresie etyki dotyczące godnej zaufania sztucznej inteligencji (Wytyczne HLEG AI), https://op.europa.eu/pl/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-pl (dostęp: 17.12.2024 r.), s. 25. Zob. też H. Felzmann, E. Fosch-Villaronga, A. Tamò-Larrieux, *Towards Transparency by Design for Artificial Intelligence*, „Science and Engineering Ethics" 2020/26, https://doi.org/10.1007/s11948-020-00276-4 (dostęp: 17.12.2024 r.), s. 3333–3361; R. Bieda, D. Lubasz, *Narzędzia wykorzystujące sztuczną inteligencję* [w:] *Analiza ryzyka i bezpieczeństwo danych w kancelariach prawnych*,

projektowania (*data protection by design*). Kluczowe znaczenie dla tej oceny może mieć również charakter tych instrumentów, nakierowanych na proaktywną i prewencyjną ochronę, integrację ochrony z celami organizacyjnymi, analizę uwzględniającą kontekst wykorzystywania i związane z tym ryzyko dla praw i wolności osób fizycznych z pogłębioną analizą wpływu, iteracyjną implementację mogącą odzwierciedlać dynamikę i wreszcie ciągłe monitorowanie w całym cyklu życia.

Badanie obejmuje weryfikację skuteczności regulacyjnej RODO w kontekście zasad godnej zaufania AI i ma dać asumpt do sformułowania wniosków z oceny adekwatności dotychczasowych ram prawnych, ewentualnie konieczności działań regulatorów, ale także potencjału osiągnięcia konwergencji poziomu ochronnego poza Unią Europejską. W tym obszarze badawczym brak jest szczegółowych opracowań zarówno w Polsce, jak i Unii Europejskiej. W literaturze prezentowane są wprawdzie wypowiedzi na temat wyzwań związanych z projektowaniem systemów AI w odniesieniu do RODO czy szerzej AI, a także przyjętego AIA. Wypowiedzi te związane są również z poszukiwaniem rozwiązań optymalizujących skuteczność regulacji, zwłaszcza w kwestii pewnego rozczarowania efektywnością RODO w obszarze cyfrowym. Nie eksplorują one jednak tematyki użyteczności modeli takich jak *data protection by design* z perspektywy zasad godnej zaufania AI i równocześnie związanych z tym wyzwań projektowych i wdrożeniowych[13]. Podjęcie się tej tematyki wypełnia zatem istniejącą lukę w literaturze, koncentrując się na pogłębionej analizie relacji między zasadami godnej zaufania AI a skutecznością regulacyjną RODO. Dotychczasowe opracowania nie odnoszą się w sposób całościowy do problematyki praktycznej implementacji zasad ochrony danych na etapie projektowania systemów AI oraz oceny ich faktycznej skuteczności w osiąganiu zakładanych w Unii

red. D. Lubasz, Warszawa 2022, s. 617–619; E. Magrani, P. Guedes Fernandes da Silva, *The Ethical and Legal Challenges of Recommender Systems Driven by Artificial Intelligence* [w:] *Multidisciplinary Perspectives on Artificial Intelligence and the Law*, red. H. Sousa Antunes i in., 2024, https://doi.org/10.1007/978-3-031-41264-6 (dostęp: 17.12.2024 r.), s. 156; M. Sadek, R.A. Calvo, C. Mougenot, *Designing value-sensitive AI: a critical review and recommendations for socio-technical design processes*, „AI and Ethics" 2024/4, https://doi.org/10.1007/s43681-023-00373-7 (dostęp: 17.12.2024 r.), s. 949–967.

[13] Szczegółowo zostało to omówione w rozdziale I pkt 4.

Europejskiej standardów ochrony. Niniejsze badanie ma zatem znaczenie zarówno dla rozwoju teorii prawa, jak i dla praktyki wdrażania regulacji w dynamicznie rozwijającym się obszarze AI.

3. Cel badań i pytania badawcze

Podtytuł monografii *Zgodność z zasadami godnej zaufania sztucznej inteligencji w modelu data protection by design* wskazuje na jej cel badawczy, który odnosi się do odpowiedzi na pytanie o możliwość wykorzystania przy projektowaniu i wdrażaniu systemów AI uregulowanego w RODO modelu *data protection by design* do osiągnięcia zgodności z zasadami godnej zaufania sztucznej inteligencji sformułowanymi przez HLEG AI.

Celem publikacji jest równocześnie opracowanie wniosków *de lege lata* i ewentualnie *de lege ferenda* oraz dostarczenie wskazówek nie tylko dla prawodawcy, regulatora, ale i dla podmiotów tworzących i wykorzystujących systemy AI, odnoszących się do wykorzystania ram regulacyjnych RODO w celu urzeczywistnienia wytycznych HLEG AI.

Osiągnięcie założonego celu badawczego wymaga odpowiedzi na szczegółowe pytania w trzech podstawowych obszarach badawczych, dotyczących: po pierwsze zasadniczej możliwości zastosowania RODO do regulowania obszaru AI, po drugie weryfikacji aplikowalności ogólnych wymogów RODO do systemów AI w kontekście wytycznych HLEG AI i wyzwań z tym związanych, po trzecie operacjonalizacji tych wymogów dzięki przewidzianym w RODO instrumentom procesowym, w szczególności *data protection by design*.

W pierwszym obszarze badawczym poszukiwana jest odpowiedź na pytania o możliwość zastosowania RODO jako elementu ram prawnych w Unii Europejskiej dla kształtowania godnej zaufania humanocentrycznej AI, a także o jego relację do AIA. W drugim obszarze znajduje się odpowiedź na pytania, w jaki sposób poszczególne wymogi wynikające z RODO mogą zostać zastosowane do projektowania i wdrażania systemów AI, a następnie czy ich konstrukcja normatywna pozwala na uwzględnienie zasad wynikających z wytycznych HLEG AI. Trzeci obszar obejmuje badanie w celu poszukiwania odpowiedzi na pytanie

o skuteczność instrumentów operacjonalizacji zgodności z wymogami RODO i wytycznymi HLEG AI.

Zakres badania, jego cel i odpowiedzi na szczegółowe pytania badawcze umożliwiają rozpoznanie nie tylko potencjału, ale i ograniczeń weryfikowanego modelu *data protection by design* z perspektywy realizacji zasad godnej zaufania AI.

4. Struktura monografii

Monografia została podzielona na trzy rozdziały, które w sposób systematyczny odpowiadają na postawione pytania badawcze. W rozdziale I przedstawiono pojęcie i wymogi odnoszące się do godnej zaufania AI, analizując definicje i ewolucję AI. W ramach tego rozdziału poruszono również kwestie etyczne oraz prawne, które mają kluczowe znaczenie dla rozwoju AI, takie jak centralizacja władzy technologicznej, wyzwania etyczne oraz rola regulacji prawnych. Następnie omówiono ramy regulacyjne AI w Unii Europejskiej, ze szczególnym uwzględnieniem genezy unijnych regulacji, zasad godnej zaufania AI oraz założeń AIA, zwracając uwagę na jego relację z RODO. Podsumowaniem rozdziału jest ocena RODO pod kątem ogólnego potencjału realizacji zasad godnej zaufania AI.

W rozdziale II skoncentrowano się na podstawowych wymogach RODO oraz ich zastosowaniu w projektowaniu i wdrażaniu systemów AI zgodnych z zasadami godnej zaufania AI. Szczególna uwaga została poświęcona podejściu opartemu na ryzyku, szacowaniu ryzyka oraz adekwatności środków organizacyjnych i technicznych w odniesieniu do ochrony praw i wolności podmiotów danych. Rozdział zawiera także analizę roli i złożonego problemu kwalifikacji podmiotów zaangażowanych w całym cyklu życia systemów AI w świetle przepisów RODO oraz szczegółowe omówienie zasad ochrony danych osobowych, takich jak legalność, rzetelność, przejrzystość, ograniczenie celu, minimalizacja danych, prawidłowość, ograniczenie przechowywania, poufność i integralność czy zasada rozliczalności. Oceniono ich zasięg normatywny i aplikowalność w celu zapewnienia realizacji zasad godnej zaufania AI. W ramach tego rozdziału przedstawiono również podstawy legalizujące

przetwarzanie danych osobowych i prawa podmiotów danych, zwracając uwagę na implikacje wynikające z profilowania i zakazu zautomatyzowanego podejmowania decyzji, zwłaszcza w kontekście nadzoru człowieka nad systemami AI.

Rozdział III dotyczy zagadnienia operacjonalizacji wymogów RODO w celu spełnienia zasad godnej zaufania AI. Skupiono się w nim na podejściu proaktywnym, mechanizmach zgodności oraz roli modelu *data protection by design*, analizując poszczególne etapy projektowania i wdrażania systemów AI – od budowania wiedzy i określania wymagań po testowanie i monitorowanie, uwzględniając aspekty praktycznego wykorzystania i wskazując przykłady konkretnych rozwiązań. Dodatkowo rozdział zawiera analizę znaczenia oceny skutków dla ochrony danych (*data protection impact assessment*, DPIA) oraz roli inspektorów ochrony danych przy wdrażaniu systemów AI. Szczególną uwagę poświęcono także realizacji zasady rozliczalności.

W ostatniej części monografii zawarto wnioski płynące z przeprowadzonych badań, analizując efektywność RODO w zapewnianiu zgodności z zasadami godnej zaufania AI oraz identyfikując ograniczenia wynikające z zakresu i charakteru regulacji. Wskazano również kierunki dalszego rozwoju, rekomendacje dla prawodawców oraz potencjalną potrzebę rewizji RODO z perspektywy wyzwań technologicznych i etycznych związanych z AI.

5. Podziękowania

Niniejsza monografia *RODO dla AI. Zgodność z zasadami godnej zaufania sztucznej inteligencji w modelu data protection by design* jest owocem lat pracy, badań oraz refleksji nad prawnymi i etycznymi wyzwaniami współczesnej technologii. Jej powstanie nie byłoby jednak możliwe bez wsparcia, inspiracji i życzliwości wielu osób, którym chciałbym w tym miejscu złożyć serdeczne podziękowania.

Na początku pragnę wyrazić swoją głęboką wdzięczność recenzentom wydawniczym, którzy poświęcili swój czas i uwagę na analizę mojej pracy. Ich konstruktywne uwagi, wnikliwe spostrzeżenia i cenne sugestie

przyczyniły się do ostatecznego kształtu niniejszej publikacji. Dziękuję za ich wkład i zaangażowanie, które pomogły mi spojrzeć na badane zagadnienia z nowych perspektyw.

Szczególne podziękowania kieruję także do moich mentorów i nauczycieli, którzy na różnych etapach mojej kariery naukowej wskazywali mi kierunki rozwoju i inspirowali do podejmowania nowych wyzwań badawczych. Ich mądrość, wiedza i wsparcie były dla mnie nieocenionym źródłem motywacji i nauki. To dzięki ich wskazówkom udało mi się zgłębić złożoność tematyki ochrony danych osobowych i jej zastosowania w kontekście AI.

Chciałbym również podziękować ekspertom w dziedzinie ochrony danych osobowych i AI, naukowcom i praktykom, z którymi miałem okazję wymieniać poglądy także podczas dyskusji panelowych i konferencji naukowych. Ich doświadczenie i wiedza były dla mnie bezcennym źródłem nowych perspektyw, inspiracji i motywacji do pogłębiania badań nad wieloaspektowością tematyki poruszanej w opracowaniu.

Pragnę z całego serca podziękować mojej rodzinie i przyjaciołom, którzy przez cały czas trwania tego projektu okazywali mi ogromną cierpliwość i wsparcie. Ich obecność była dla mnie oparciem w chwilach trudności. Jestem wdzięczny za ich miłość, przyjaźń, wyrozumiałość i gotowość do wysłuchania, nawet gdy rozmowy o RODO i AI zdawały się nie mieć końca.

Nie sposób nie wspomnieć o moich wspólnikach i współpracownikach w kancelarii, których dodatkowe zaangażowanie i pomoc w codziennych obowiązkach zawodowych umożliwiły mi skupienie się na pracy naukowej. Ich gotowość do przejęcia części moich zadań była nieocenioną pomocą i pozwoliła mi w pełni poświęcić się realizacji tego projektu. Ich wsparcie było nie tylko praktyczne, ale także stanowiło źródło motywacji w trudnych momentach.

Na koniec pragnę podziękować wszystkim, którzy w jakikolwiek sposób przyczynili się do powstania tej monografii – czy to poprzez inspirację, dyskusję, wsparcie techniczne, czy też po prostu dobrą radę. Każdy z Was

wniósł swoją cegiełkę do realizacji tego przedsięwzięcia, za co jestem niezmiernie wdzięczny.

Mam nadzieję, że niniejsza publikacja będzie nie tylko przyczynkiem do dalszych dyskusji naukowych, ale także praktycznym narzędziem dla tych, którzy na co dzień mierzą się z wyzwaniami ochrony danych osobowych związanymi z AI.

Część badań, zwłaszcza odnoszących się do zagadnień technologicznych, etycznych oraz podejścia opartego na wartościach (*values by design*), które ma szeroki potencjał wykorzystania z uwagi na możliwości ogniskowania ochrony, w szczególności na różnych prawach podstawowych, została częściowo sfinansowania przez Narodowe Centrum Nauki w ramach projektu badawczego „Ochrona konsumenta a sztuczna inteligencja. Między prawem a etyką" (nr projektu: 2018/31/B/HS5/01169), wybranego w konkursie OPUS 16, którego jestem jednym z wykonawców.

Rozdział I

GODNA ZAUFANIA SZTUCZNA INTELIGENCJA

1. Ewolucja i pojęcie sztucznej inteligencji

1.1. Ewolucja sztucznej inteligencji

Historia AI rozpoczęła się od podstawowych pytań filozoficznych odnoszących się do natury ludzkiego myślenia i możliwości jego symulacji przez maszyny. Jednym z kluczowych momentów było wprowadzenie przez A. Turinga w 1950 r. tzw. testu Turinga, mającego na celu ocenę zdolności maszyn do symulacji ludzkiego zachowania[1]. W swoim arty-

[1] M. Haenlein i A. Kaplan sięgają jednak dalej do lat 40. XX w., poszukując korzeni dziedziny nauki odnoszącej się do sztucznej inteligencji. Nawiązują oni do opublikowanego w 1942 r. opowiadania I. Asimova *Zabawa w berka* (*Runaround*), w którym autor sformułował trzy prawa robotyki, tj.: 1) robot nie może zranić człowieka ani poprzez bezczynność pozwolić człowiekowi na wyrządzenie krzywdy; 2) robot musi być posłuszny rozkazom wydawanym mu przez ludzi, z wyjątkiem sytuacji, gdy takie rozkazy byłyby sprzeczne z pierwszym prawem; 3) robot musi chronić swoje własne istnienie, o ile taka ochrona nie jest sprzeczna z pierwszym lub drugim prawem. Praca Asimowa inspirowała bowiem pokolenia naukowców w dziedzinie robotyki, sztucznej inteligencji i informatyki, w tym amerykańskiego kognitywistę M.L. Minsky'ego (późniejszego współzałożyciela laboratorium sztucznej inteligencji Massachusetts Institute of Technology), któremu, wraz z J. McCarthy, N. Rochesterem i C. Shannonem, przypisuje się sformułowanie pojęcia „sztuczna inteligencja", użytego w tytule projektu badawczego *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, zainicjowanego 31.08.1955 r. (zob. J. McCarthy, M.L. Minsky, N. Rochester, C.E. Shannon, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955*, „AI Magazine"

kule *Computing Machinery and Intelligence*[2] zaproponował on odejście od filozoficznych spekulacji na rzecz praktycznej analizy, czy maszyna może w przekonujący sposób imitować zachowanie ludzkie w ramach tzw. gry naśladownictwa (*imitation game*). Ten przełomowy koncept położył fundament pod interdyscyplinarne badania łączące matematykę, filozofię, informatykę i neurobiologię.

Formalne uznanie AI za odrębną dziedzinę nauki nastąpiło podczas konferencji w Dartmouth w 1956 r., która zgromadziła takich pionierów jak J. McCarthy, M.L. Minsky, N. Rochester i C. Shannon[3]. Celem nowej dziedziny miało być stworzenie maszyn zdolnych do rozwiązywania problemów, uczenia się i rozumowania w sposób przypominający działania ludzkiego umysłu. W ramach wczesnych badań wykształciły się dwa główne nurty – symboliczne przetwarzanie informacji oraz symulacja biologiczna.

Symboliczne przetwarzanie informacji, inspirowane logiką formalną, koncentrowało się na manipulowaniu abstrakcyjnymi symbolami zgodnie z ustalonymi regułami. Jednym z pierwszych przykładów tego podejścia był system Logic Theorist, opracowany przez laureata Nagrody Nobla H. Simona i naukowców z RAND Corporation C. Shawa i A. Newella[4], który automatycznie rozwiązywał twierdzenia matematyczne. Równolegle rozwijano symulację biologiczną, naśladującą procesy zachodzące w ludzkim mózgu. W tym nurcie A. Newell

2006/27(4), https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/1904 (dostęp: 18.12.2024 r.), s. 12–14; zob. szerzej pkt 1.2). Dla A. Turinga natomiast jednym ze źródeł refleksji na temat natury ludzkiego myślenia był m.in. proces opracowywania przez niego maszyny do łamania kodów o nazwie The Bombe, powszechnie uważanej za pierwszy działający komputer elektromechaniczny. Zob. M. Haenlein, A. Kaplan, *A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence*, „California Management Review" 2019/61(4), https://doi.org/10.1177/0008125619864925 (dostęp: 17.12.2024 r.), s. 2.

[2] A. Turing, *Computing Machinery and Intelligence*, „Mind" 1950/59(236), s. 433–460.

[3] Projekt badawczy *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, zainicjowany 31.08.1955 r. Zob. J. McCarthy, M.L. Minsky, N. Rochester, C.E. Shannon, *A Proposal...*, s. 12–14.

[4] A. Toosi, A. Bottino, B. Saboury, E. Siegel, A. Rahmim, *A Brief History of AI: How to Prevent Another Winter (A Critical Review)*, „PET Clinics" 2021/16(4), s. 6.

i H. Simon opracowali program GPS (*General Problem Solver*), który został zaprojektowany tak, aby naśladować protokoły rozwiązywania problemów przez ludzki mózg[5]. Jest on uważany za pierwszą pracę w ramach „ludzkiego rozumowania" AI. Przełomowym osiągnięciem był tu jednak dopiero Mark I – perceptron zbudowany w Cornell w 1957 r. przez F. Rosenblatta, uważanego za ojca głębokiego uczenia się[6]. Został on skonstruowany w oparciu o analogową jednowarstwową sieć neuronową zdolną do klasyfikowania danych wejściowych na dwie potencjalne kategorie, z możliwością uczenia się metodą prób i błędów[7]. Perceptron Mark I jest uważany za jednego z przodków współczesnych sieci neuronowych[8].

Mniej więcej w tym samym czasie A. Samuel prowadził, na przykładzie systemu do gry w warcaby, badania nad uczeniem się ze wzmocnieniem, tj. rodzajem algorytmu AI, w którym model AI uczy się, jak wchodzić w interakcje z otaczającym go środowiskiem, aby osiągnąć swój cel za pomocą systemu opartego na nagrodach[9]. Jego pracę uznaje się za pierwszy przykład programu AI oparty na uczeniu się ze wzmocnieniem i, jak się wskazuje, nawet za przodka późniejszych systemów, takich jak TD-GAMMON z 1992 r., jednego z najlepszych na świecie graczy

[5] A. Newell, J. Shaw, H.A. Simon, *Report on a general problem solving program* [w:] *Proceedings of the International Conference on Information Processing*, Paris 1959, s. 64; M. Haenlein, A. Kaplan, *A Brief...*, s. 3–4.

[6] C.C. Tappert, *Who is the father of deep learning?* [w:] *2019 International Conference on Computational Science and Computational Intelligence (CSCI)*, https://www.researchgate.net/publication/340810637_Who_Is_the_Father_of_Deep_Learning (dostęp: 18.12.2024 r.), s. 343–348.

[7] F. Rosenblatt, *Report No. 85-460-1 The perceptron, a perceiving and recognizing automaton (Project para)*, Cornell Aeronautical Laboratory, Buffalo, New York 1957, https://bpb-us-e2.wpmucdn.com/websites.umass.edu/dist/a/27637/files/2016/03/rosenblatt-1957.pdf (dostęp: 18.12.2024 r.).

[8] M. Lefkowitz, *Professor's perceptron paved the way for AI – 60 years too soon*, 25.09.2019, https://news.cornell.edu/stories/2019/09/professors-perceptron-paved-way-ai-60-years-too-soon (dostęp: 18.12.2024 r.).

[9] A.L. Samuel, *Some studies in machine learning using the game of checkers*, „IBM Journal of Research and Development" 1959/3(3), https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5392560 (dostęp: 18.12.2024 r.), s. 211–229.

w tryktraka (*backgammon*) i AlphaGo z 2016 r., który pokonał ludzkiego mistrza świata w Go[10].

Po konferencji w Dartmouth nastąpił okres prawie dwóch dekad, nazwanych wiosną lub latem AI, w którym odnotowano znaczący rozwój w dziedzinie AI[11]. W tym czasie oprócz wyżej wspomnianych powstał m.in. program komputerowy ELIZA, stworzony w latach 1964–1966 przez J. Weizenbauma na Massachusetts Institute of Technology. Był on narzędziem do przetwarzania języka naturalnego, zdolnym do symulowania rozmowy z człowiekiem i jednym z pierwszych programów, które były w stanie przejść wspomniany wcześniej test Turinga[12].

W 1958 r. J. McCarthy z kolei wprowadził specyficzny dla AI język programowania Lisp, który przez następne trzy dekady był dominującym językiem programowania AI[13], kładąc również podwaliny pod koncepcyjne podejście do systemów AI, oparte na reprezentacji wiedzy i rozumowaniu[14]. W tym samym roku pierwsze eksperymentalne prace obejmujące algorytmy ewolucyjne w AI zostały przeprowadzone przez R.M. Freidberga w kierunku genetycznego programowania[15]. Natomiast w 1959 r. N. Rochester i H. Gelernter z IBM opracowali program do sprawdzania twierdzeń geometrycznych napisany w języku Fortran[16]. W latach 60. XX w. M.L. Minsky zaproponował po-

[10] A. Toosi, A. Bottino, B. Saboury, E. Siegel, A. Rahmim, *A Brief*..., s. 7; D. Silver i in., *Mastering the game of Go with deep neural networks and tree search*, „Nature" 2016/529, https://www.nature.com/articles/nature16961 (dostęp: 18.12.2024 r.), s. 484–489.

[11] A. Toosi, A. Bottino, B. Saboury, E. Siegel, A. Rahmim, *A Brief*..., s. 5; M. Haenlein, A. Kaplan, *A Brief*..., s. 1 i n.

[12] https://www.masswerk.at/elizabot/ (dostęp: 18.12.2024 r.).

[13] A. Toosi, A. Bottino, B. Saboury, E. Siegel, A. Rahmim, *A Brief*..., s. 7.

[14] J. McCarthy, *Programs with common sense*, 1959, http://jmc.stanford.edu/articles/mcc59/mcc59.pdf (dostęp: 18.12.2024 r.).

[15] R.M. Friedberg, *A learning machine: Part I*, „IBM Journal of Research and Development" 1958/2(1), https://ieeexplore.ieee.org/document/5392654 (dostęp: 18.12.2024 r.), s. 2–13; S. Russell, P. Norvig, *Sztuczna inteligencja. Nowe spojrzenie*, Gliwice 2023, s. 39.

[16] H. Gelernter, J.R. Hansen, C.L. Gerberich, *A Fortran-Compiled List-Processing Language*, „Journal of the Association for Computing Machinery", 7.04.1960, https://dl.acm.org/doi/pdf/10.1145/321021.321022 (dostęp: 18.12.2024 r.); H. Gelernter, J.R. Hansen, D.W. Loveland, *Empirical explorations of the geometry theorem machine* [w:] *Proceedings of the Western Joint IRE-AIEE-ACM Computer Conference, San Francisco, CA, USA, 3–5*

dejście upraszczające dla przypadków użycia AI tzw. mikroświatów (*microworlds*)[17], które stało się podstawą dla dalszych prac, m.in. programu SAINT J. Slagle'a z 1963 r.[18], programu ANALOGY T.G. Evansa z 1964 r.[19] oraz programu STUDENT, napisanego w języku Lisp przez D.G. Bobrowa w 1967 r.[20]

W 1970 r. na fali entuzjazmu związanego z szybko postępującemu rozwojem AI M.L. Minsky udzielił wywiadu dla „Life Magazine", w którym stwierdził, że maszyna o ogólnej inteligencji przeciętnego człowieka może zostać opracowana w ciągu trzech do ośmiu lat, co, jak wiemy, do dziś nie zostało osiągnięte i jest nadal w fazie rozważań teoretycznych. Zaledwie trzy lata później Kongres Stanów Zjednoczonych zakwestionował wysokie wydatki na badania nad AI na podstawie raportu ALPAC z 1966 r.[21] Z kolei w roku 1972 matematyk J. Lighthill opublikował raport zlecony przez British Science Research Council, który zawierał analogicznie negatywne wnioski, krytykujące optymistyczne prognozy badaczy AI[22]. O ograniczeniach, zwłaszcza w obszarze rozwoju AI z wykorzystaniem sieci neuronowych, pozwalających na wykonywanie

May 1960, https://dl.acm.org/doi/pdf/10.1145/1460361.1460381 (dostęp: 18.12.2024 r.), s. 143–149.

[17] A. Toosi, A. Bottino, B. Saboury, E. Siegel, A. Rahmim, *A Brief*..., s. 7; S. Russell, P. Norvig, *Sztuczna inteligencja*..., s. 38.

[18] Program ten służył do rozwiązywania zamkniętych problemów całkowania w rachunku różniczkowym. Zob. J.R. Slagle, *A Heuristic Program that Solves Symbolic Integration Problems in Freshman Calculus*, „Journal of the Association for Computing Machinery" 1963/10(4), https://dl.acm.org/doi/pdf/10.1145/321186.321193 (dostęp: 18.12.2024 r.), s. 507–520.

[19] Program ten służył do rozwiązywania problemów geometrycznych w teście IQ. Zob. T.G. Evans, *A Heuristic Program to Solve Geometric-Analogy Problems* [w:] *Proceedings of the 1964 spring joint computer conference*, AFIPS 1964, http://logical.ai/auai/p327-evans.pdf (dostęp: 18.12.2024 r.), s. 327–338.

[20] Program ten służył do rozwiązywania zadań z algebry na podstawie poleceń wydawanych w języku naturalnym. Zob. D.G. Bobrow, *Natural Language Input for a Computer Problem Solving System*, Massachusetts 1964, https://www.researchgate.net/publication/37597683_Natural_Language_Input_for_a_Computer_Problem_Solving_System (dostęp: 18.12.2024 r.).

[21] A. Toosi, A. Bottino, B. Saboury, E. Siegel, A. Rahmim, *A Brief*..., s. 8.

[22] J. Lighthill, *Artificial Intelligence: A General Survey*, http://www.chilton-computing.org.uk/inf/literature/reports/lighthill_report/p001.htm (dostęp: 18.12.2024 r.); M. Haenlein, A. Kaplan, *A Brief*..., s. 3.

niesformalizowanych zadań, np. na bazie rozpoznawania obrazów, pisali również M.L. Minsky i S. Paperta w *Perceptrons: An Introduction to Computational Geometry*[23]. W publikacji tej wskazali na fundamentalne ograniczenia sieci neuronowych, związane przede wszystkim z brakiem wystarczającej mocy obliczeniowej ówczesnych komputerów.

Ostatecznie na wiele lat rozwój badań w tej dziedzinie został zahamowany i w latach 70. XX w. entuzjazm wobec AI osłabł, co doprowadziło do okresu stagnacji określanego mianem „pierwszej zimy AI" (*first AI winter*)[24].

W latach 70 i 80. XX w. następowały krótkie fazy ożywienia dzięki systemom ekspertowym[25], które wykorzystywały reguły *if-then* w celu rozwiązywania problemów m.in. w dziedzinie medycyny[26]. Okresy te przeplatały się ze spowolnieniem wynikającym z niskiej użyteczności, zbyt wąskich zastosowań, niewystarczającego stopnia elastyczności i wysokich kosztów utrzymania systemów ekspertowych.

Historia rozwoju AI nie jest zatem liniowym ciągiem sukcesów technologicznych, lecz obejmuje również okresy stagnacji. Zjawisko to występowało wielokrotnie w drugiej połowie XX w. i, generalizując, wynikało z nadmiernych oczekiwań wobec AI, które w zestawieniu z technicznymi ograniczeniami tamtych czasów oraz brakiem odpowiedniego wsparcia

[23] M.L. Minsky, S.A. Papert, *Perceptrons: An Introduction to Computational Geometry*, Cambridge 1969.

[24] A. Toosi, A. Bottino, B. Saboury, E. Siegel, A. Rahmim, *A Brief*..., s. 8; M. Haenlein, A. Kaplan, *A Brief*..., s. 3–4.

[25] Są one zwane również systemami eksperckimi. Zob. S. Russell, P. Norvig, *Sztuczna inteligencja*..., s. 40–42.

[26] Przykładami mogą tu być system DENDRAL, stworzony w Stanford przez E.A. Feigenbauma, B.G. Buchanana i J. Lederberga, którzy wnioskowali o strukturze molekularnej na podstawie danych spektrometrii mas (E.A. Feigenbaum, B.G. Buchanan, J. Lederberg, *On generality and problem solving: A case study using the DENDRAL program*, 1970, https://www.researchgate.net/publication/23865744_On_generality_and_problem_solving_A_case_study_using_the_DENDRAL_program [dostęp: 18.12.2024 r.]) czy system MYCIN, opracowany przez Edwarda H. Shortliffe'a i B.G. Buchanana, służący do diagnozowania zakażeń krwi (E.H. Shortliffe, B.G. Buchanan, *A model of inexact reasoning in medicine*, „Mathematical Biosciences" 1975/23, https://stacks.stanford.edu/file/druid:ts764ph5106/ts764ph5106.pdf [dostęp: 18.12.2024 r.], s. 351–379).

finansowego prowadziły do zniechęcenia środowisk naukowych, inwestorów i opinii publicznej. Okresy kryzysów miały jednak pozytywny wpływ na rozwój AI. W tym czasie badacze skoncentrowali się na rozwoju podstaw teoretycznych oraz ulepszaniu technologii sprzętowych, takich jak procesory graficzne (*graphics processing unit*, GPU). Te postępy stworzyły fundamenty dla późniejszych przełomów w dziedzinie uczenia maszynowego[27].

Dzięki ponownie uruchomionym w latach 80. XX w. inwestycjom, m.in. powołaniu przez rząd USA Microelectronics and Computer Technology Corporation, ożywiono badania, które doprowadziły w latach 90. XX w. i na początku XXI w. do przełomu w postaci uczenia maszynowego (*machine learning*). W klasycznej AI dominowało podejście symboliczne, oparte na logice i manipulacji strukturami językowymi. Systemy te wymagały manualnego tworzenia reguł i nie radziły sobie z dynamicznymi środowiskami, w których pojawiają się dane niepełne, niepewne lub zmienne. Rewolucja uczenia maszynowego nastąpiła, gdy badacze zaczęli włączać elementy statystyki i teorii prawdopodobieństwa w proces projektowania algorytmów. Ich filarami stało się wykorzystanie algorytmów optymalizacyjnych, tj. w szczególności takich technik, jak regresja logistyczna, drzewa decyzyjne czy początkowe wersje sieci neuronowych, co umożliwiło automatyczne dostosowywanie parametrów modelu w celu maksymalizacji jego wydajności oraz modelowanie probabilistyczne z wykorzystaniem przykładowo takich algorytmów jak naiwny klasyfikator bayesowski (*Naive Bayes Classifier*) czy ukryte modele Markowa (*Hidden Marcov Models*, HMMs)[28], które wprowadziły mechanizmy wnioskowania statystycznego, pozwalając maszynom ra-

[27] Między innymi badania K. Fukushimy dotyczące głębokich konwolucyjnych sieci neuronowych (*convolutional neural networks*, CNN). Zob. K. Fukushima, *Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position*, „Biological Cybernetics" 1980/36, https://www.rctn.org/bruno/public/papers/Fukushima1980.pdf (dostęp: 18.12.2024 r.), s. 193–202.

[28] L.E. Baum, T. Petrie, *Statistical Inference for Probabilistic Functions of Finite State Markov Chains*, „The Annals of Mathematical Statistics" 1966/37, https://projecteuclid.org/journals/annals-of-mathematical-statistics/volume-37/issue-6/Statistical-Inference-for-Probabilistic-Functions-of-Finite-State-Markov-Chains/10.1214/aoms/1177699147.full (dostęp: 18.12.2024 r.), s. 1554–1563; S. Russell, P. Norvig, *Sztuczna inteligencja*..., s. 43.

dzić sobie z niepewnością i brakami w danych. W metodach tych zatem odchodzi się od podejścia symbolicznego na rzecz metod opartych na danych. Umożliwiło to systemom uczenie się na podstawie danych i dostosowywanie swoich działań w dynamicznych środowiskach. Kluczowym wydarzeniem było zwycięstwo programu IBM Deep Blue nad mistrzem świata w szachach G. Kasparowem w 1997 r., symbolizujące potencjał nowych technik.

Przełom w uczeniu maszynowym byłby niemożliwy bez postępu technologicznego w zakresie mocy obliczeniowej[29] oraz wielkoskalowego przetwarzania i przechowywania danych (*big data*)[30]. Rosnąca ilość danych cyfrowych, generowanych przez użytkowników internetu, systemy IoT oraz urządzenia mobilne stworzyły niezbędne zasoby do trenowania modeli uczenia maszynowego. Równocześnie upowszechnienie chmury obliczeniowej (*cloud computing*) i wprowadzenie platform, takich jak Amazon Web Services (AWS) i Google Cloud, uczyniło wysokowydajne obliczenia bardziej dostępnymi dla szerokiego grona badaczy i przedsiębiorstw. Obraz dopełniło opracowanie procesorów graficznych, które umożliwiały wykonywanie obliczeń równoległych, co znacząco zwiększyło efektywność trenowania złożonych modeli, m.in. głębokich sieci neuronowych.

Rozwój technologii obliczeniowej i dostępność danych pozwoliły na powrót do koncepcji sieci neuronowych jako algorytmów maszynowego uczenia. Wprawdzie zespół Y. LeCun już w 1989 r. podjął badania nad konwolucyjnymi sieciami neuronowymi (*convolutional neural network*, CNN)[31], jednak dopiero w 2012 r. nastąpił przełom jakościowy

[29] Mowa tu o procesorach graficznych (GPU), tensorowych (TPU), ale także programowalnych macierzach bramek (FPGA). Zob. S. Russell, P. Norvig, *Sztuczna inteligencja*..., s. 45.

[30] M.A. Morris, B. Saboury, B. Burkett, J. Gao, E.L. Siegel, *Reinventing Radiology: Big Data and the Future of Medical Imaging*, „Journal of Thoracic Imaging" 2018/33, https://journals.lww.com/thoracicimaging/abstract/2018/01000/reinventing_radiology__big_data_and_the_future_of.3.aspx (dostęp: 18.12.2024 r.), s. 4–16.

[31] Y. LeCun i in., *Backpropagation Applied to Handwritten Zip Code Recognition*, „Neural Computation" 1989/1(4), https://yann.lecun.com/exdb/publis/pdf/lecun-89e.pdf (dostęp: 18.12.2024 r.), s. 541–551.

w grupie badawczej G. Hintona na Uniwersytecie w Toronto w ramach ILSVRC (*ImageNet Large Scale Visual Recognition Challenge*)[32]. Jest to również związane z kolejną fazą rozwoju uczenia maszynowego, znaną jako głębokie uczenie (*deep learning*). Opracowanie i wdrożenie tych algorytmów uczenia maszynowego uważane jest za początek obecnego etapu dynamicznego rozwoju AI o szerokim spektrum zastosowań.

Technologia ta wykorzystuje najczęściej wielowarstwowe sieci neuronowe do wykrywania złożonych wzorców w danych. Ich przykładem może być AlexNet – głęboka konwolucyjna sieć neuronowa, której architekturę zaproponował A. Krizhevsky, czy ResNet (*Residual Network*) – również rodzaj konwolucyjnej sieci neuronowej, zaprojektowanej w celu rozwiązania problemu degradacji w głębokich sieciach neuronowych. Kluczowymi, świadczącymi o olbrzymim potencjale szerokiego zastosowania, cechami głębokiego uczenia są automatyczne wykrywanie cech (*feature extraction*), wykorzystywanie algorytmów wstecznej propagacji (*backpropagation*), co pozwala modelom dostosowywać wagi sieci na podstawie błędów popełnionych w procesie uczenia i wreszcie wykorzystanie wielkich zbiorów danych treningowych obejmujących potencjalnie miliardy przykładów.

Kolejnym przełomem w erze głębokiego uczenia się było opracowanie wielkoskalowych modeli językowych (*large language models*, LLM), takich jak GPT (*Generative Pre-trained Transformer*) czy BERT (*Bidirectional Encoder Representations from Transformers*). Modele te, oparte na architekturze transformatorowej, zrewolucjonizowały dziedziny związane z przetwarzaniem języka naturalnego (*natural language processing*, NLP). Są one przykładami tzw. generatywnych modeli AI (*generative AI models*, GenAI), tj. modeli uczenia maszynowego, które uczą się rozkładu probabilistycznego danych wejściowych i potrafią na tej podstawie tworzyć zasadniczo nowe dane, podobne do danych treningowych. W przeciwieństwie do modeli dyskryminacyjnych, opartych zasadniczo na klasyfikacji, które skupiają się na klasyfikacji lub przewidywaniu etykiet na

[32] A. Toosi, A. Bottino, B. Saboury, E. Siegel, A. Rahmim, *A Brief*..., s. 5; M. Haenlein, A. Kaplan, *A Brief*..., s. 1 i n. Na temat samego konkursu zob. https://www.image-net.org/challenges/LSVRC/ (dostęp: 18.12.2024 r.).

podstawie cech danych wejściowych, generatywne modele koncentrują się na symulacji i tworzeniu danych, takich jak teksty (np. GPT-3, GPT-4 i kolejne), obrazy (np. DALL·E, Stable Diffusion), dźwięki (np. OpenAI Jukebox) i wideo (np. Synthesia, StyleGAN). Reprezentowane są tu różne modele maszynowego uczenia, nie tylko wspomniany odnośnie do LLM model transformatorowy, ale również m.in. autoenkodery wariacyjne (*variational autoencoders*, VAEs)[33], generatywne sieci adwersarialne (*generative adversarial networks*, GANs)[34] czy modele dyfuzji (*diffusion models*)[35]. Charakterystyczne dla tych modeli jest ich pretrenowanie na dużych zbiorach danych, możliwość multimodalności (np. GPT-4, Gemini i Claude 3) oraz uniwersalność zastosowań – od tworzenia nowych treści, personalizacji, poprzez optymalizacje procesów biznesowych, edukację, a skończywszy na diagnostyce i leczeniu.

Uwzględniając potencjał rozwoju technologii w obszarze AI, szczególnie zaawansowanych modeli, w 2023 r. wprowadzono termin *Frontier AI*. W lipcu 2023 r. takie organizacje, jak Anthropic, Google, Microsoft i OpenAI ogłosiły utworzenie Frontier Model Forum[36], mającego na celu promowanie bezpiecznego i odpowiedzialnego rozwoju zaawansowanych systemów AI. W tym samym miesiącu OpenAI opublikowało dokument Frontier AI regulation: Managing emerging risks to public safety[37], w którym termin *Frontier AI* odnosi się do wysoko zaawansowanych modeli AI o potencjalnie niebezpiecznych zdolnościach. W sposób bardziej uogólniony *Frontier AI* to pojęcie odnoszące się do najbardziej zaawansowanych i innowacyjnych form AI, które mają potencjał wprowadzać przełomowe zmiany w technologii, społeczeństwie i gospodarce. Jest ono często używane w kontekście technologii znajdujących się na

[33] Opracowane przez D.P. Kingmę i M. Wellinga. Zob. D.P. Kingma, M. Welling, *Auto-Encoding Variational Bayes*, https://arxiv.org/abs/1312.6114 (dostęp: 18.12.2024 r.).

[34] Modele GAN zaproponowane przez I.J. Goodfellowa w 2014 r. Zob. I.J. Goodfellow i in., *Generative Adversarial Nets*, https://proceedings.neurips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf (dostęp: 18.12.2024 r.).

[35] Zob. Z. Chang, G. Koulieris, H.P.H. Shum, *On the Design Fundamentals of Diffusion Models: A Survey*, https://arxiv.org/pdf/2306.04542 (dostęp: 18.12.2024 r.).

[36] https://www.frontiermodelforum.org (dostęp: 18.12.2024 r.).

[37] M. Anderljung i in., *Frontier AI regulation: Managing emerging risks to public safety*, 6.07.2023 r., https://openai.com/index/frontier-ai-regulation/ (dostęp: 18.12.2024 r.).

granicy obecnych możliwości badawczych i inżynieryjnych, a także w odniesieniu do przyszłościowych systemów, które mogą znacząco wpłynąć na sposób, w jaki ludzie współdziałają z technologią w różnych aspektach życia. Jednocześnie *Frontier AI* stawia przed ludzkością wyzwania związane z bezpieczeństwem, etyką i zrównoważonym rozwojem, które muszą być odpowiednio adresowane. Jego cechami charakterystycznymi są potencjał transformacyjny, skalowalność oraz autonomiczność i wreszcie szerokie spektrum dziedzin, w których może mieć zastosowanie, w tym w medycynie (np. diagnostyka oparta na głębokim uczeniu), przemyśle zbrojeniowym (autonomiczne drony bojowe), edukacji (personalizowane ścieżki uczenia), a także w badaniach naukowych i eksploracji kosmosu.

Nie podejmując się dokonywania predykcji dalszego rozwoju systemów AI, warto jednak zwrócić uwagę na generalne trendy i prawa, którym podlega rozwój technologii.

W pierwszej kolejności można sięgnąć do corocznych badań prowadzonych w ramach projektu *One Hundred Year Study on AI* (AI100) na Uniwersytecie Stanforda[38] oraz podprojektu AI Index, ustanowionego pod parasolem AI100 i obecnie realizowanego w Stanford Institute for Human-Centered Artificial Intelligence (HAI). AI Index to próba śledzenia, zestawiania, destylacji i wizualizacji danych dotyczących AI. Aspiruje do bycia kompleksowym źródłem danych i analiz dla decydentów, badaczy, kadry kierowniczej, dziennikarzy i ogółu społeczeństwa, aby rozwijać intuicje na temat złożonej AI[39].

W najnowszym raporcie podkreślono obecnie kluczowe znaczenie dużych modeli językowych, których liczba na całym świecie w 2023 r. podwoiła się w porównaniu z rokiem poprzednim. Model Gemini Ultra, opracowany przez Google DeepMind (DeepMind Technologies Limited), stał się pierwszym modelem LLM, który osiągnął wydajność na poziomie ludzkim w teście porównawczym *Massive Multitask Language*

[38] https://ai100.stanford.edu (dostęp: 18.12.2024 r.).
[39] https://ai100.stanford.edu/ai-index (dostęp: 18.12.2024 r.).

Understanding (MMLU)[40]. W raporcie podkreślono, że generalnie wyniki systemów AI przewyższają ludzkie w kilku testach porównawczych, m.in. w klasyfikacji obrazów, rozumowaniu wizualnym i rozumieniu języka angielskiego. Pozostają jednak w tyle w bardziej złożonych zadaniach, takich jak matematyka na poziomie konkursowym, wizualne rozumowanie i planowanie[41].

Według raportu Gartnera za 2023 r.[42] w horyzoncie czasowym do dwóch lat rozwiną się technologie takie jak *edge AI*[43], rozpoznawanie obrazu, także widzenie komuperowe (*computer vision*)[44], do pięciu lat: *causal AI*[45], *first-principles AI* (FPAI), także *physics-informed AI*[46], grafy wiedzy (*knowledge graphs*)[47], złożona AI (*composite AI*)[48], a do 10 lat: modele

[40] D. Hendrycks i in., *Measuring Massive Multitask Language Understanding* [w:] *International Conference on Learning Representations, ICLR 2021*, https://arxiv.org/pdf/2009.03300 (dostęp: 18.12.2024 r.).

[41] https://aiindex.stanford.edu/report/ (dostęp: 18.12.2024 r.).

[42] Zob. *What's New in Artificial Intelligence from the 2023 Gartner Hype Cycle*, https://www.gartner.com/en/articles/what-s-new-in-artificial-intelligence-from-the-2023-gartner-hype-cycle (dostęp: 18.12.2024 r.).

[43] Edge AI odnosi się do wykorzystania technik AI osadzonych w produktach innych niż IT, punktach końcowych IoT, bramach i serwerach brzegowych. Obejmuje przypadki użycia w zastosowaniach konsumenckich, komercyjnych i przemysłowych, takich jak autonomiczne pojazdy, zwiększone możliwości diagnostyki medycznej i strumieniowa analiza wideo.

[44] *Computer vision* to zestaw technologii, które obejmują przechwytywanie, przetwarzanie i analizowanie rzeczywistych obrazów i filmów w celu wyodrębnienia znaczących, kontekstowych informacji ze świata fizycznego.

[45] *Casual AI* to przyczynowa AI, która identyfikuje i wykorzystuje związki przyczynowo-skutkowe, aby wyjść poza modele predykcyjne oparte na korelacji i dążyć do systemów AI, które mogą skuteczniej zalecać działania i działać bardziej autonomicznie.

[46] *First-principles AI* włącza zasady fizyczne i analogowe, prawa rządzące i wiedzę dziedzinową do modeli AI. Rozszerza inżynierię AI na inżynierię złożonych systemów i systemy oparte na modelach.

[47] *Knowledge graphs* to nadające się do odczytu maszynowego reprezentacje świata fizycznego i cyfrowego. Obejmują one podmioty (osoby, firmy, zasoby cyfrowe) i ich relacje, które są zgodne z modelem danych grafu.

[48] *Composite AI* odnosi się do połączonego zastosowania (lub fuzji) różnych technik AI w celu poprawy wydajności uczenia się i poszerzenia poziomu reprezentacji wiedzy. Rozwiązuje to szerszy zakres problemów biznesowych w bardziej efektywny sposób.

ogólnego przeznaczenia (*foundation models*)[49], systemy wieloagentowe (*multiagent systems*, MAS)[50], systemy autonomiczne (*autonomic systems*)[51], operacyjne systemy AI (*operational AI systems*, OAISys)[52] czy inteligentne roboty (*smart robots*)[53].

Już jednak w raporcie za 2024 r.[54] dostrzegalne są zasadnicze zmiany w projekcji, nie tylko w zakresie etapu cyklu popularności (*hype cycle*)[55], ale i szybkości osiągnięcia dojrzałości. Najszybciej rozwijającymi się technologiami lub procesami technologicznymi, które powinny osiągnąć swoją dojrzałość, czyli tzw. płaskowyż produktywności, tj. fazę rozwoju, w której wzrost zaczyna zwalniać, ale generowane są znaczne przychody i produkt zaczyna być traktowany jako coś oczywistego[56], stają się *composite AI*, *knowledge graphs* i *AI engineering*. Autorzy raportu wskazują, że wspomniany *composite AI* stanowi kolejną fazę ewolucji AI[57].

---

[49] *Foundation modells* to modele o dużych parametrach, trenowane na szerokiej gamie zestawów danych w sposób samonadzorowany.

[50] *Multiagent systems* (MAS) to rodzaj systemu AI składającego się z wielu niezależnych (ale interaktywnych) agentów, z których każdy może postrzegać swoje środowisko i podejmować działania. Agentami mogą być modele AI, programy komputerowe, roboty i inne jednostki obliczeniowe.

[51] *Autonomic systems* to samozarządzające się systemy fizyczne lub programowe, wykonujące zadania związane z domeną, które wykazują trzy podstawowe cechy: autonomię, uczenie się i sprawczość.

[52] *Operational AI systems* (OAISys) umożliwiają orkiestrację, automatyzację i skalowanie gotowej do produkcji AI klasy korporacyjnej, obejmującej ML, DNN i generatywną AI.

[53] *Smart robots* to napędzane AI, często mobilne maszyny zaprojektowane do autonomicznego wykonywania jednego lub więcej zadań fizycznych.

[54] Zob. A. Jaffri, *Explore Beyond GenAI on the 2024 Hype Cycle for Artificial Intelligence*, https://www.gartner.com/en/articles/hype-cycle-for-artificial-intelligence (dostęp: 18.12.2024 r.).

[55] *Hype cycle* według Gartnera składa się z rozłożonych na osi oczekiwań względem osi czasu etapów: 1) trigger innowacji, 2) szczyt rozbudzonych oczekiwań, 3) rynna rozczarowania, 4) zbocze oświecenia i 5) płaskowyż produktywności.

[56] A. Jaffri, *Explore Beyond GenAI...*, *passim*.

[57] Złożona AI odnosi się do połączonego zastosowania (lub fuzji) różnych technik AI w celu poprawy efektywności uczenia się i poszerzenia poziomu reprezentacji wiedzy. Poszerza ona mechanizmy abstrakcji AI i ostatecznie zapewnia platformę do rozwiązywania szerszego zakresu problemów biznesowych w bardziej efektywny sposób. Zob. https://www.gartner.com/en/information-technology/glossary/composite-ai (dostęp: 18.12.2024 r.).

Takie podejście umożliwia organizacjom maksymalizację wpływu ich inicjatyw dotyczących AI, prowadząc do dokładniejszych prognoz, decyzji i automatyzacji, nawet w złożonych środowiskach. Z kolei *AI engineering* jest podstawą do dostarczania AI i GenAI na skalę korporacyjną[58]. Większości organizacji brakuje danych, analiz i podstaw oprogramowania, aby przenieść poszczególne projekty AI do produkcji na dużą skalę, a tym bardziej obsługiwać portfolio takich projektów. Podejścia inżynieryjne AI, takie jak DataOps, ModelOps i DevOps, umożliwiają natomiast wdrażanie modeli do produkcji w ustrukturyzowanej, powtarzalnej strukturze modelu fabrycznego. Ostatecznie grafy wiedzy (*knowledge graphs*) to czytelne dla maszyn reprezentacje świata fizycznego i cyfrowego. Przechwytują informacje w intuicyjnym wizualnie formacie, a jednocześnie są w stanie reprezentować złożone relacje. Zapewniają one niezawodną logikę i możliwe do wyjaśnienia rozumowanie, w przeciwieństwie do zawodnych, ale potężnych możliwości predykcyjnych GenAI[59].

Na te trendy należy jeszcze nałożyć tendencje inwestycyjne na świecie[60]. Stany Zjednoczone nadal prowadzą w rankingu Global AI Index 2024 (62,5 mld USD). W 2024 r. powiększyły swoją przewagę nad Chinami (7,3 mld USD), które utrzymały drugie miejsce. Te dwa supermocarstwa znacznie wyprzedzają wszystkie inne kraje w indeksie. Dla przykładu Unia Europejska razem z Wielką Brytanią to łącznie 9 mld USD inwestycji w tworzenie i rozwijanie AI[61]. Szacowany na 2023 r. globalny rynek prywatnych inwestycji w AI wynosił 130 mld USD[62], ale Statista

[58] Inżynieria AI (*AI engineering*) – związek między sztuczną inteligencją a wartością biznesową. Jest to dyscyplina służąca integracji podejść DataOps, MLOps i DevOps w celu zapewnienia spójnego opracowania, dostarczania i wdrażania systemów opartych na AI w organizacji. Zob. https://www.gartner.com/en/webinar/573970/1290752 (dostęp: 18.12.2024 r.).

[59] A. Jaffri, *Explore Beyond GenAI...*, *passim*.

[60] The Global AI Index2024, https://www.tortoisemedia.com/intelligence/global-ai/ (dostęp: 18.12.2024 r.); szczegółowy raport, https://www.tortoisemedia.com/2024/09/19/the-global-artificial-intelligence-index-2024/ (dostęp: 18.12.2024 r.).

[61] https://www.tortoisemedia.com/intelligence/global-ai/ (dostęp: 18.12.2024 r.).

[62] https://www.europarl.europa.eu/RegData/etudes/ATAG/2024/760392/EPRS_ATA(2024)760392_EN.pdf (dostęp: 18.12.2024 r.).

szacuje, że do roku 2030 rynek ten osiągnie blisko 1,9 tryliona USD[63]. Równocześnie globalny rynek generatywnej AI jest w dużej mierze zdominowany przez Stany Zjednoczone, a konkretnie przez kilku gigantów technologicznych. Tym niemniej według Gartnera GenAI w 2024 r. przekroczyła szczyt zawyżonych oczekiwań. Do końca 2024 r., jak twierdzą autorzy, wartość będzie w dużej mierze pochodzić z projektów opartych na znanych technikach AI, samodzielnych lub w połączeniu z GenAI, które mają ustandaryzowane procesy wspomagające wdrażanie. Zamiast skupiać się wyłącznie na GenAI, liderzy AI powinni szukać złożonych technik AI, które łączą podejścia z innowacji na wszystkich etapach cyklu popularności. Jednak GenAI nadal ma potencjał, aby stać się technologią transformacyjną o głębokim wpływie biznesowym na odkrywanie treści, tworzenie, autentyczność i regulacje, automatyzację pracy ludzkiej oraz doświadczenia klientów i pracowników. Według Gartnera sama liczba i skala projektów AI oznacza, że liderzy AI muszą rozszerzyć zakres zainteresowania AI poza kwestie techniczne. Organizacje powinny zwrócić szczególną uwagę na obszary, takie jak zarządzanie, w tym zarządzanie danymi (*data governance*), zarządzanie zmianą, odpowiedzialność, bezpieczeństwo i łagodzenie długu technicznego, jeśli chcą sprostać wyzwaniom w zakresie regulacyjnym, etyki, ale i skalowalności.

Z powołanych raportów wynika ostatecznie m.in. również to, że wspomniana wcześniej ogólna AI (*Artificial General Intelligence*, AGI) będzie możliwa do osiągnięcia dopiero za więcej niż 10 lat. Jednak jak widać na przykładzie porównania, opisanych w raportach z 2023 i 2024 r., tendencji technologicznych, wystarczył rok na zmianę priorytetów technologicznych i istotne postępy w niektórych dziedzinach względem innych prognozowanych rok wcześniej. Zmienia się także poziom inwestycji w poszczególnych państwach oraz obszarach, co istotnie oddziałuje na potencjał wzrostowy, ale nie o charakterze liniowym. Powoduje to, że długoterminowe predykcje rozwoju są bardzo utrudnione i konieczna jest ich bieżąca aktualizacja.

[63] https://www.statista.com/forecasts/1474143/global-ai-market-size (dostęp: 18.12.2024 r.).

Problem z przewidzeniem szybkości i kierunku rozwoju technologicznego jest też związany z prawami rządzącymi tym rozwojem. Należy tu wskazać choćby na prawa Moore'a i Kurzweila. Zgodnie z pierwszym, o charakterze empirycznym, ekonomicznie optymalna liczba tranzystorów w układzie scalonym zwiększa się w kolejnych latach zgodnie z trendem wykładniczym (podwaja się w niemal równych odcinkach czasu)[64]. Prawo to *per analogiam* używane jest do określenia praktycznie dowolnego postępu technologicznego.

Z kolei R. Kurzweil zdefiniował prawo przyspieszających zwrotów (*law of accelerating returns*) również odnoszące się do rozwoju technologicznego. Wskazuje na konieczność rozszerzenia prawa Moore'a, twierdząc, że „istnieje nawet wykładniczy wzrost w tempie wykładniczego wzrostu"[65]. W historii technologii przyspieszenie zmian to postrzegany wzrost tempa zmian technologicznych na przestrzeni dziejów, co może sugerować szybsze i głębsze zmiany w przyszłości i może, ale nie musi towarzyszyć im równie głęboka zmiana społeczna i kulturowa.

W swoich kolejnych predykcjach R. Kurzweil idzie jednak dalej uważając, że ewolucja dostarcza dowodów na to, iż pewnego dnia ludzie stworzą maszyny bardziej inteligentne niż oni sami. Sięga tym samym do pojęcia technicznej osobliwości (*technical singularity*)[66]. Jej osiągnięcie spowoduje, że dalszy rozwój techniczny przestanie być przez człowieka kontrolowany i stanie się nieodwracalny, co doprowadzi do głębokich i nieprzewidywalnych zmian w ludzkiej cywilizacji i nieodwracalnego

[64] Autorstwo tego prawa przypisuje się G.E. Moore'owi, jednemu z założycieli przedsiębiorstwa Intel. Zob. G.E. Moore, *Cramming more components onto integrated circuits*, http://svmoore.pbworks.com/w/file/fetch/59055901/Gordon_Moore_1965_Article.pdf (dostęp: 18.12.2024 r.).

[65] R. Kurzweil, *The Law of Accelerating Returns*, https://web.archive.org/web/20100619033859/http:/www.kurzweilai.net/articles/art0134.html?printable=1 (dostęp: 18.12.2024 r.); R. Kurzweil, *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*, New York 1999.

[66] R. Kurzweil, *The Singularity Is Near: When Humans Transcend Biology*, New York 2005, s. 23. Termin pochodzi od węgierskiego matematyka Johna von Neumanna, ale podstawy teoretyczne sięgają prac A. Turinga, zwłaszcza *Computing Machinery and Intelligence*, w których wprowadza ideę zdolności maszyny do wykazywania inteligentnego zachowania równoważnego lub nieodróżnialnego od ludzkiego.

przekształcenia ludzkiego życia[67]. Termin „osobliwość" w tym kontekście wywodzi się z koncepcji matematycznych wskazujących na punkt, w którym istniejące modele załamują się, a ciągłość rozumienia zostaje utracona. Opisuje to erę, w której maszyny nie tylko dorównują ludzkiej inteligencji, ale znacznie ją przewyższają, rozpoczynając cykl samonapędzającej się ewolucji technologicznej. Rozważania na temat sztucznej superinteligencji (*artificial superintelligence*) i jej potencjału, skłoniły niektórych badaczy, m.in. S. Hawkinga, do stawiania najdalej idących tez, że tego typu technologia jest w istocie wynalazkiem dnia ostatecznego (*The Doomsday Invention*)[68].

Sprowadzając te rozważania na poziom potrzeb regulacyjnych, należy oczywiście pamiętać o etapie rozwoju technologii tak w obszarze szczegółowym, odnoszącym się do poszczególnych zastosowań, jak i w kategorii AI. W tym drugim przypadku jesteśmy, z technologicznego punktu widzenia, w fazie wykorzystywania i projektowania słabej AI (*weak, narrow AI*). Powoduje to, że ocena wpływu zarówno na jednostkę, jak i na społeczeństwa jest bardziej skonkretyzowana zastosowaniami. Z perspektywy jednak potencjału rozwoju do poziomu ogólnej AI i teoretycznie również superinteligencji kluczowa staje się odpowiedź na bardzo podstawowe, wręcz egzystencjalne pytania. Będzie tu miało bowiem niebagatelne znaczenie zdefiniowanie zasad, którymi ma rządzić się nie tyle sama technologia, co jej rozwój[69]. Wykracza to jednak daleko poza ramy tego opracowania, w którym skupiono się na aplikacji dostępnych obecnie lub w niedługiej przyszłości technologii w oparciu o stworzone ramy etyczne. Rozważania te, uwzględniając humanocentryczne podejście, nie pozostają jednak bez wpływu na rozwój technologii i mechanizmy, których możliwe, a wręcz konieczne będzie

[67] R. Kurzweil, *The Singularity*..., s. 5. Swoje rozważania na ten temat autor kontynuuje w książce *Singularity is nearer*, wskazując na przyspieszające tempo zmian. Zob. R. Kurzweil, *Osobliwość coraz bliżej*, Kielce 2024.

[68] *Stephen Hawking*..., *passim*; zob. też K. Roose, *A.I. Poses*..., *passim*; W. Hartzog, *Two AI Truths*..., s. 612; M. Suleyman, M. Bhaskar, *Nadchodząca fala. Sztuczna inteligencja, władza i najważniejszy dylemat ludzkości XXI wieku*, Kraków 2024, s. 22–23.

[69] J. Schuett, *Defining the scope of AI regulations*, „Law, Innovation and Technology" 2021/4, LawAI Working Paper No. 4-2021, https://ssrn.com/abstract=3453632 (dostęp: 18.12.2024 r.), s. 60–82.

zaimplementowanie w procesie tworzenia nowych rozwiązań, niezależnie od tego jak prawdopodobne jest ich stworzenie[70].

1.2. Definicja sztucznej inteligencji

Rozwój AI stanowi jedno z kluczowych osiągnięć współczesnej nauki i technologii, wpływając na różne aspekty życia społecznego, gospodarczego i politycznego. Ewolucja AI przebiegała przez kolejne etapy, począwszy od początkowych koncepcji teoretycznych, przez rozwój technik uczenia maszynowego i tworzenie systemów ekspertowych, aż po współczesny boom związany z głębokim uczeniem (*deep learning*) i wielkoskalowymi modelami językowymi. Ta dynamiczna ewolucja, zwłaszcza w ostatnim okresie, ze względu na rozwój *big data* i poprawę mocy obliczeniowej, przyniosła liczne osiągnięcia, ale także nowe wyzwania technologiczne, prawne i społeczne.

Interesujące jest występujące w dyskursie na temat AI zjawisko, że na potrzeby dyskusji często się jej nie definiuje[71]. Możliwą przyczyną takiego stanu rzeczy jest zdiagnozowany przez E. Yudkowskiego syndrom, wskazujący, że zdecydowanie największym niebezpieczeństwem AI jest to, że ludzie zbyt wcześnie dochodzą do wniosku, że ją rozumieją[72]. Jego konsekwencją jest założenie, iż wszyscy uczestnicy debaty rozumieją niezdefiniowane pojęcie AI w taki sam sposób, co oczywiście jest założeniem błędnym. Dalszym następstwem jest zjawisko nazywane *AI*

[70] V.C. Müller, N. Bostrom, *Future Progress in Artificial Intelligence: A Survey of Expert Opinion* [w:] *Fundamental Issues of Artificial Intelligence*, red. V.C. Müller, Cham 2016, https://doi.org/10.1007/978-3-319-26485-1_33 (dostęp: 17.12.2024 r.); S.D. Baum, B. Goertzel, T.G. Goertzel, *How Long Until Human-Level AI? Results from an Expert Assessment*, „Technological Forecasting and Social Change" 2011/78(1), https://doi.org/10.1016/j.techfore.2010.09.006 (dostęp: 17.12.2024 r.), s. 185 i n.; J. Schuett, *Defining*..., s. 60–82.

[71] T. Zalewski, *Definicja sztucznej inteligencji* [w:] *Prawo sztucznej inteligencji*, red. L. Lai, M. Świerczyński, Warszawa 2020, s. 1 i n.

[72] E. Yudkowski, *Artificial Intelligence as a Positive and Negative Factor Global Risk* [w:] *Global Catastrophic Risks*, red. N. Bostrom, M.M. Ćirković, New York 2008, https://intelligence.org/files/AIPosNegFactor.pdf (dostęp: 18.12.2024 r.), s. 1: „By far, the greatest danger of Artificial Intelligence is that people conclude too early that they understand it".

washing, które oznacza praktykę marketingową polegającą na takim oznakowaniu produktów lub usług, by sprawiały wrażenie wykorzystujących AI[73]. W rzeczywistości rozwiązania te albo oferują zwykłą automatyzację procesów z wykorzystaniem algorytmów, podstawowe operacje statystyczne, albo są po prostu istniejącymi już aplikacjami, jedynie promowanymi z zastosowaniem nowego chwytu marketingowego[74].

Z tych przyczyn dla ukierunkowania dyskursu niezbędne jest wskazanie przynajmniej kluczowych cech tego zjawiska, aby w ujęciu teleologicznym można było określić przedmiot i oczekiwane efekty regulacyjne. W tym celu niezbędne jest określenie podstawowych cech kwalifikacyjnych AI lub co najmniej jej aplikacji, czyli systemów AI. Należy zatem sięgnąć do podstawowych, występujących w nauce, definicji AI lub systemów AI, dostrzegając przy tym, że pierwsze z tych pojęć odnosi się do dziedziny nauki, a drugie do jej aplikacji w postaci systemu o określonych cechach, obie jednak powinny odnosić się do podstawowych właściwości funkcjonalnych technologii.

Definicja AI ewoluuje od lat 50. XX w. i wywodzi się wprost z analiz A. Turinga i tzw. testu Turinga, dziś bardziej znanego pod skrótem CAPTCHA (*Completely Automated Public Turing test to tell Computers and Humans Apart*). Zgodnie z tą definicją „sztuczna inteligencja" to zdolność maszyny do naśladowania lub imitowania ludzkiej inteligencji[75]. Dla celów dalszej analizy ta definicja nie jest jednak przydatna, podobnie jak uważana za jedną z pierwszych definicja sformułowana przez J. McCarthy'ego, M.L. Minsky'ego, N. Rochestera i C.E. Shannona w ramach projektu *Dartmouth Summer Research Project on Artificial*

[73] T. Grzegory, J. Puskas, *LegalTech, czyli jak branża prawnicza wkracza w „dwudziesty" wiek* [w:] *LegalTech, czyli jak bezpiecznie korzystać z narzędzi informatycznych w organizacji, w tym w kancelarii oraz dziale prawnym*, red. D. Szostek, Warszawa 2021, s. 565.

[74] T. Grzegory, J. Puskas, *LegalTech*..., s. 565; S. Overby, *AI vs. automation: 6 ways to spot fake AI*, https://enterprisersproject.com/article/2020/3/ai-vs-automation-6-ways-spot-fake-ai (dostęp: 18.12.2024 r.).

[75] J. Schuett, *A Legal Definition of AI*, https://www.researchgate.net/publication/335600149_A_Legal_Definition_of_AI (dostęp: 18.12.2024 r.), s. 3; T. Zalewski, *Definicja*..., s. 1.

Intelligence[76]. Zgodnie z ich propozycją „sztuczna inteligencja" to techniczne rozwiązanie wykonujące czynności będące zazwyczaj domeną ludzi, szczególnie wymagające użycia ludzkiego intelektu. Sztuczna inteligencja to zatem maszyna, która zachowuje się tak jak człowiek, maszyna, która myśli. W podobny sposób kognitywista M.L. Minsky uznał sztuczną inteligencję za „naukę o sprawianiu, by maszyny robiły rzeczy, które wymagałyby inteligencji, gdyby były wykonywane przez ludzi"[77]. Definicje te mają charakter socjologiczny i subiektywny, odnoszący się do oceny osoby rozróżniającej AI od nie-AI[78]. Nie wskazują one natomiast elementów dystynktywnych tego pojęcia, z którymi wiążą się określone skutki, ale także z którymi można powiązać oczekiwania normatywne.

Dla celów regulacyjnych przy definiowaniu AI należy uwzględnić perspektywę funkcjonalną, ale także techniczną i technologiczną. Formułując relewantne technologicznie i funkcjonalnie cechy, w pierwszej kolejności trzeba wskazać, jak słusznie zauważa T. Zalewski, że elementem konstytutywnym każdej inteligencji, w tym AI, jest zdolność do uczenia się i zdolność do samodzielnego rozwiązywania problemów[79]. W odniesieniu do systemów AI chodzi o kompetencję do uczenia się na podstawie dostarczonych zbiorów danych, z wykorzystaniem algorytmów uczenia maszynowego (*machine learning*)[80], np. regresji liniowej i logistycznej, drzew decyzyjnych, sieci neuronowych (*artificial neural networks*, ANN), głębokiego uczenia (*deep learning*) czy hierarchicznych metod klasteryzacji[81]. W zależności od wybranej metody uczenia system jest w stanie generować różne wyniki odnoszące się do klasyfikacji, re-

[76] Uznaje się ich również za twórców pojęcia „sztuczna inteligencja" (*artificial intelligence*), którego użyli w tytule swojego projektu badawczego *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, zainicjowanego 31.08.1955 r. Zob. J. McCarthy, M.L. Minsky, N. Rochester, C.E. Shannon, *A Proposal*..., s. 12–14.

[77] M.L. Minsky, *Semantic information processing*, Cambridge 1969.

[78] Na temat charakteru definicji Turinga zob. szerzej T. Zalewski, *Definicja*..., s. 1.

[79] T. Zalewski, *Definicja*..., s. 4.

[80] Uczenie maszynowe to dziedzina nauki odnosząca się do zdolności systemów do nauki bez konieczności ich jawnego programowania. Zob. A.L. Samuel, *Some studies*..., s. 211–229.

[81] Poszczególne z tych algorytmów wykorzystywane są przy różnych metodach uczenia maszynowego, tj. uczenia nadzorowango (*supervised learning*), nienadzorowa-

gresji, korelacji itd. W ujęciu generalnym podczas procesu trenowania modelu AI z wykorzystaniem przyjętego algorytmu analizowane są zatem dostarczone dane treningowe w celu wykrycia wzorców i relacji, które pozwolą następnie wykonywać określone różne zadania, np. klasyfikować obrazy, analizować teksty i prognozować trendy. System AI powinien równocześnie wykazywać pewne, zależne od konfiguracji, zdolności adaptacyjne i działanie autonomiczne oraz wchodzić w interakcję z otoczeniem[82].

Przy formułowaniu definicji AI problematyczne jest to, że technologia ta ma bardzo dynamiczny charakter[83], co z jednej strony wpływa na zmienność jej szczegółowych cech, a z drugiej na potencjał oddziaływania. Dlatego w doktrynie co do zakresu znaczeniowego pojęcia „sztuczna inteligencja" trwa ożywiona dyskusja[84]. Można jednak pokusić się o przytoczenie najogólniejszej definicji, by ukierunkować dalsze rozważania. Zgodnie z nią „sztuczna inteligencja" definiowana jest jako dziedzina nauki zajmująca się tworzeniem systemów posiadających zdolność do prawidłowej interpretacji danych zewnętrznych, uczenia się na podstawie tych danych i wykorzystywania sformułowanych na tej

nego (*unsupervised learning*) czy ze wzmocnieniem (*reinforcement learning*). Zob. szerzej rozdział III pkt 2.3.5.

[82] Na te cechy zwraca uwagę T. Zalewski, *Definicja*..., s. 2–3.

[83] Zob. pkt 1.1. Zob. też UNESCO, Recommendation on the Ethics of Artificial Intelligence, 23.11.2021, https://unesdoc.unesco.org/ark:/48223/pf0000381137 (dostęp: 18.12.2024 r.), s. 10.

[84] Zob. M. Legg, F. Bell, *Artificial Intelligence and the Legal Profession: Becoming The AI-Enhanced Lawyer*, „University of Tasmania Law Review" 2019/38(2), s. 38; T. Grzegory, J. Puskas, *LegalTech*..., s. 565 i n.; J. Schuett, *Defining*..., s. 60–82; M. Haenlein, A. Kaplan, *Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence*, „Business Horizons" 2018/62, https://www.researchgate.net/publication/328761767_Siri_Siri_in_my_hand_Who's_the_fairest_in_the_land_On_the_interpretations_illustrations_and_implications_of_artificial_intelligence (dostęp: 18.12.2024 r.); M. Świerczyński, *AI a praca prawnika w świetle wytycznych Rady Europy* [w:] *LegalTech, czyli jak bezpiecznie korzystać z narzędzi informatycznych w organizacji, w tym w kancelarii oraz dziale prawnym*, red. D. Szostek, Warszawa 2021, s. 520; P. Vogel, *Künstliche Intelligenz und Datenschutz*, Baden-Baden 2022, s. 34; T. Zalewski, *Definicja*..., s. 2 i n.

podstawie wniosków do osiągania określonych celów i realizacji zadań poprzez elastyczną adaptację[85].

W zbliżonym kierunku poszły analizy regulacyjne oraz definiowanie AI czy też częściej systemów AI jako konkretnego rozwiązania technicznego w kolejnych gremiach międzynarodowych.

Definicję AI znajdziemy w Europejskiej karcie etycznej w sprawie wykorzystania sztucznej inteligencji w systemach sądowych i ich otoczeniu, przyjętej przez Europejską Komisję ds. Skuteczności Wymiaru Sprawiedliwości[86]. Zgodnie z nią pojęcie „sztuczna inteligencja" odnosi się do zbioru metod, teorii i technik naukowych, których celem jest odtworzenie przez maszynę zdolności poznawczych człowieka[87].

Rada Europy także zdefiniowała pojęcie systemów AI w Konwencji ramowej w sprawie sztucznej inteligencji, praw człowieka, demokracji i praworządności[88]. Zgodnie z art. 3 konwencji „system sztucznej inteligencji" oznacza system, który dla celów jawnych lub ukrytych wnioskuje, na podstawie otrzymanych danych wejściowych, w jaki sposób wygenerować wyniki, takie jak prognozy, treści, zalecenia lub decyzje, które mogą wpływać na środowisko fizyczne lub wirtualne. Poszczególne systemy AI różnią się między sobą poziomem autonomii i zdolności adaptacyjnych po wdrożeniu.

---

[85] M. Haenlein, A. Kaplan, *A Brief*..., s. 1.

[86] European Commission for the Efficiency of Justice (CEPEJ), European Ethical Charter on the use of Artificial Intelligence in judicial systems and their environment, Adopted at the 31st plenary meeting of the CEPEJ, Strasbourg, 3–4 December 2018, https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c (dostęp: 27.12.2024 r.).

[87] European Commission for the Efficiency of Justice (CEPEJ), European Ethical Charter on the use of Artificial Intelligence in judicial systems and their environment, Adopted at the 31st plenary meeting of the CEPEJ, Strasbourg, 3–4 December 2018, https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c (dostęp: 27.12.2024 r.), s. 69.

[88] Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, Vilnius, 5.09.2024, https://rm.coe.int/1680afae3c (dostęp: 18.12.2024 r.).

Warta odnotowania jest również wypracowana w drodze konsensusu w grupie ekspertów w ramach rekomendacji dla krajów OECD i członków stowarzyszonych definicja przyjęta przez OECD w zaleceniach Rady w sprawie sztucznej inteligencji z 2019 r.[89] Określono w nich, że „system sztucznej inteligencji" to system oparty na maszynie, który dla celów jawnych lub niejawnych wnioskuje na podstawie otrzymanych danych wejściowych, w jaki sposób generować dane wyjściowe, takie jak prognozy, treści, zalecenia lub decyzje, które mogą wpływać na środowisko fizyczne lub wirtualne, przy czym różne systemy AI różnią się poziomem autonomii i zdolności adaptacyjnych po wdrożeniu.

W Unii Europejskiej zadanie sformułowania lub weryfikacji definicji „sztucznej inteligencji" Komisja Europejska powierzyła HLEG AI w ramach zadania o szerszym zakresie, tj. opracowania wytycznych w zakresie etyki dotyczących godnej zaufania sztucznej inteligencji oraz zaleceń na temat polityki i inwestycji[90]. W towarzyszącym wytycznym dokumencie Definicja SI. Główne funkcje i dyscypliny[91] HLEG AI wskazała, że za punkt wyjścia przyjęła definicję zaproponowaną przez Komisję w komunikacie w sprawie sztucznej inteligencji[92], zgodnie z którą termin „sztuczna inteligencja" odnosi się do systemów, które

[89] Recommendation of the OECD Council on Artificial Intelligence, 22.05.2019, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 (dostęp: 18.12.2024 r.).

[90] Grupa ekspertów wysokiego szczebla ds. sztucznej inteligencji (HLEG AI), Wytyczne w zakresie etyki dotyczące godnej zaufania sztucznej inteligencji, https://op.europa.eu/pl/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-pl (dostęp: 3.04.2024 r.). Wytyczne HLEG AI legły również u podstaw wniosku legislacyjnego Komisji dotyczącego rozporządzenia Parlamentu Europejskiego i Rady ustanawiającego zharmonizowane przepisy dotyczące sztucznej inteligencji (Akt w sprawie sztucznej inteligencji) i zmieniającego niektóre akty ustawodawcze Unii, 21.04.2021, COM(2021) 206 final; zob. uzasadnienie wniosku, https://eur-lex.europa.eu/legal-content/PL/TXT/?uri=CELEX%3A52021PC0206#footnote24 (dostęp: 17.12.2024 r.); zob. uzasadnienie wniosku, pkt 3.2.

[91] Grupa ekspertów wysokiego szczebla ds. sztucznej inteligencji (HLEG AI), Definicja SI. Główne funkcje i dyscypliny, https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60670 (dostęp: 18.12.2024 r.).

[92] Komunikat Komisji do Parlamentu Europejskiego, Rady Europejskiej, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów, Sztuczna inteligencja dla Europy, 25.04.2018 r., COM(2018) 237 final.

wykazują inteligentne zachowanie dzięki analizie otoczenia i podejmowaniu działań – do pewnego stopnia autonomicznie – w celu osiągnięcia konkretnych celów. Systemy AI mogą być oparte na oprogramowaniu, działając w świecie wirtualnym (np. asystenci głosowi, oprogramowanie do analizy obrazu, wyszukiwarki, systemy rozpoznawania mowy i twarzy), lub mogą być wbudowane w urządzenia (np. zaawansowane roboty, samochody autonomiczne, drony lub aplikacje internetu rzeczy).

Dokonując jednak analizy stanu wiedzy aktualnej na rok 2018 oraz wyodrębniając cechy charakterystyczne systemów AI, takie jak uczenie się[93], rozumowanie i podejmowanie decyzji na bazie danych z otoczenia[94] i wreszcie interakcje ze środowiskiem zewnętrznym, HLEG AI zaproponowała zrewidowaną definicję. Zgodnie z nią „systemy sztucznej inteligencji" to oprogramowanie komputerowe (i ewentualnie również sprzęt komputerowy) stworzone przez człowieka, które, biorąc pod uwagę złożony cel, działają w wymiarze fizycznym lub cyfrowym poprzez postrzeganie ich otoczenia dzięki gromadzeniu danych, interpretacji zebranych ustrukturyzowanych lub nieustrukturyzowanych danych, rozumowaniu na podstawie wiedzy lub przetwarzaniu informacji pochodzących z tych danych oraz podejmowaniu decyzji w sprawie najlepszych działań, które należy podjąć w celu osiągnięcia określonego celu. Systemy AI mogą wykorzystywać symboliczne reguły albo uczyć się modelu numerycznego, a także dostosowywać swoje zachowanie, analizując wpływ ich poprzednich działań na otoczenie. Jako dyscyplina naukowa AI obejmuje rożne podejścia i techniki, takie jak uczenie się maszyn (czego konkretnymi przykładami są uczenie głębokie i uczenie przez wzmacnianie), rozumowanie maszyn (obejmujące planowanie, programowanie działań, reprezentowanie wiedzy i rozumowanie, wyszukiwanie i optymalizację) oraz robotyka (obejmująca sterowanie,

[93] Uczenie maszynowe z wykorzystaniem dostępnych metod, z których najbardziej rozpowszechnione są uczenie nadzorowane, uczenie nienadzorowane i uczenie przez wzmacnianie, wykorzystujące odpowiednie dla nich algorytmy. Zob. wyżej w tym punkcie.

[94] Ta grupa technik, jak wskazuje HLEG AI, obejmuje reprezentowanie wiedzy i rozumowanie, planowanie, programowanie działań, wyszukiwanie i optymalizację. Techniki te umożliwiają prowadzenie rozumowania w odniesieniu do danych pochodzących z czujników. Zob. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (dostęp: 18.12.2024 r.).

postrzeganie, czujniki i urządzenia wykonawcze, a także integrację wszystkich innych technik w systemach cyberfizycznych).

Definicja ta stała się podstawą definicji systemu AI zaproponowanej przez Komisję Europejską we wniosku legislacyjnym dotyczącym rozporządzenia Parlamentu Europejskiego i Rady ustanawiającego zharmonizowane przepisy dotyczące sztucznej inteligencji (Akt w sprawie sztucznej inteligencji)[95]. W toku prac legislacyjnych zdecydowano się jednak uelastycznić proponowaną definicję i uczynić ją bardziej neutralną technologicznie, korzystając z dorobku OECD. Ostatecznie przyjęto w art. 3 pkt 1 AIA, że „system sztucznej inteligencji" oznacza system maszynowy, który został zaprojektowany do działania z różnym poziomem autonomii po jego wdrożeniu oraz który może wykazywać zdolność adaptacji po jego wdrożeniu, a także który – na potrzeby wyraźnych lub dorozumianych celów – wnioskuje, jak generować na podstawie otrzymanych danych wejściowych wyniki, takie jak predykcje, treści, zalecenia lub decyzje, które mogą wpływać na środowisko fizyczne lub wirtualne.

Nieco odmienne podejście przyjęło natomiast UNESCO, proponując by nie definiować pojęcia „system sztucznej inteligencji", a skupić się na jego cechach, które mają istotne znaczenie etyczne[96]. W związku tym wyróżniono trzy elementy, które zajmują centralne miejsce. Po pierwsze system AI to zdolność uczenia się i wykonywania zadań poznawczych prowadzących do wyników takich jak przewidywanie i podejmowanie decyzji w środowiskach materialnych i wirtualnych. W tym kontekście systemy AI są projektowane do działania z różnym stopniem autonomii za pomocą modelowania i reprezentacji wiedzy oraz przez wykorzystywanie danych i obliczanie korelacji. Po wtóre systemy AI są dynamiczne i stanowią wyzwania etyczne we wszystkich etapach cyklu życia, począwszy od badań, projektowania i rozwoju, przez wdrożenie

[95] Wniosek dotyczący rozporządzenia Parlamentu Europejskiego i Rady ustanawiającego zharmonizowane przepisy dotyczące sztucznej inteligencji (Akt w sprawie sztucznej inteligencji) i zmieniającego niektóre akty ustawodawcze Unii, 21.04.2021, COM(2021) 206 final.

[96] UNESCO, Recommendation on the Ethics of Artificial Intelligence, 23.11.2021, https://unesdoc.unesco.org/ark:/48223/pf0000381137 (dostęp: 18.12.2024 r.), s. 10.

i użytkowanie, w tym utrzymanie, eksploatację, handel, finansowanie, monitorowanie i ocenę, walidację, po zakończenie użytkowania, demontaż i zakończenie. Po trzecie systemy AI generują nowe rodzaje kwestii etycznych, które obejmują m.in. ich wpływ na podejmowanie decyzji, zatrudnienie i pracę, interakcje społeczne, opiekę zdrowotną, edukację, media, dostęp do informacji, przepaść cyfrową, dane osobowe i ochronę konsumentów, środowisko, demokrację, praworządność, bezpieczeństwo i policję, podwójne zastosowanie oraz prawa człowieka i podstawowe wolności, w tym wolność słowa, prywatność i niedyskryminację. Co więcej, nowe wyzwania etyczne powstają w związku z potencjałem systemów AI powielania i wzmacniania istniejących zniekształceń, a tym samym zaostrzania już istniejących form dyskryminacji, uprzedzeń i stereotypów[97].

To ostatnie podejście jest mi najbliższe z uwagi z jednej strony na jego neutralność technologiczną, z drugiej zaś na odniesienie do dystynktywnych cech systemów AI, z którymi związana jest możliwość oddziaływania tak na jednostki, jak i na całe społeczeństwa, co ostatecznie wiąże się wyzwaniami regulacyjnymi.

Dla celów dalszej analizy kluczowymi elementami, wskazywanymi również w powyższych definicjach i wyjaśnieniach, będą zdolność uczenia się, pewien stopień autonomii przy przewidywaniu i podejmowaniu decyzji, możliwości adaptacji i oddziaływanie na świat zewnętrzny, fizyczny lub cyfrowy w oparciu o wnioskowanie na podstawie danych wejściowych.

Terminem relewantnym dla wywodów będzie zatem nie tyle „sztuczna inteligencja", ale „system sztucznej inteligencji" jako konkretny, zbudowany system lub aplikacja, która implementuje techniki AI, aby realizować określone zadania. System taki może być oprogramowaniem, urządzeniem fizycznym lub ich kombinacją. Odnosi się to do wdrożenia praktycznych rozwiązań AI, które działają w rzeczywistych środowiskach fizycznych lub cyfrowych i mają określone funkcje. Dopiero

[97] UNESCO, Recommendation on the Ethics of Artificial Intelligence, 23.11.2021, https://unesdoc.unesco.org/ark:/48223/pf0000381137 (dostęp: 18.12.2024 r.), s. 10–11.

bowiem urzeczywistnione założenia teoretyczne mają potencjał rzeczywistego oddziaływania, również negatywnego. Ramy etyczne i prawne muszą dotyczyć w konsekwencji aplikacji AI w formie systemu. Będą wtedy bowiem oddziaływały w szczególności na kształt obowiązków producentów systemów AI i podmiotów je stosujących.

Na koniec, dla porządku, należy jeszcze zauważyć, że AI dzieli się w nauce także na słabą lub wąską (*weak AI* lub *narrow AI*), silną lub ogólną (*strong AI* lub *Artificial General Intelligence*, AGI) i superinteligencję (*Artificial Super Intelligence*, ASI) na podstawie kryterium etapu jej ewolucji[98]. Obecnie badania nad AI, jak i wywody zawarte w niniejszym opracowaniu koncentrują się głównie na dziedzinie słabej AI, której wykorzystanie ogranicza się w istocie do określonych zadań. Systemy słabej AI projektowane są do wykonywania określonych, wyspecjalizowanych zadań[99]. Słaba AI nie posiada świadomości ani zdolności ogólnego rozumowania, a jej działanie zawęża się do wąskiego obszaru, w którym została zaprogramowana. Realizujące takie zadania systemy AI są obecnie w powszechnym użytku[100]. W przyszłości natomiast możliwe jest powstanie następnej generacji AI, tj. AI ogólnej, zdolnej do rozumowania, planowania i autonomicznego rozwiązywania problemów w zadaniach, do których nigdy nie została zaprojektowana. Obecne w nauce można spotkać również wypowiedzi o możliwości opracowania w dalszej przyszłości trzeciej generacji AI, tzn. sztucznej superinteligencji (ASI), która mogłaby być samoświadoma, zdolna do kreatywności naukowej, umiejętności społecznych i ogólnej mądrości, dlatego niektórzy nazywają ją prawdziwą AI[101].

---

[98] P. Vogel, *Künstliche Intelligenz*..., s. 37–38; M. Haenlein, A. Kaplan, *Siri, Siri*..., *passim*.

[99] Odnosi się to także do systemów generatywnej AI, która należy do kategorii słabej AI, a jej zadaniem jest tworzenie nowych treści, zależnie od typu obrazów, dźwięków czy tekstu.

[100] Na przykład asystenci głosowi (np. Siri, Alexa), algorytmy rekomendacji (np. Netflix, Spotify), systemy rozpoznawania twarzy, autonomiczne pojazdy (na obecnym etapie rozwoju). Zob. szerzej M. Haenlein, A. Kaplan, *Siri, Siri*..., *passim*.

[101] M. Haenlein, A. Kaplan, *Siri, Siri*..., *passim*.

2. Etyczne i prawne aspekty w rozwoju sztucznej inteligencji

2.1. Zagadnienia podstawowe

Dynamiczny rozwój AI otworzył szerokie możliwości technologiczne, podwyższając pułap potencjału i ambicji człowieka[102]. Wykorzystanie technologii w kierunku uwalniania człowieka od powtarzalnych zadań przy równoczesnym wzroście efektywności pracy twórczej, zwiększeniu skuteczności i szybkości badań naukowych, zwłaszcza w dziedzinie medycyny czy zmian klimatu, ma niebagatelne znaczenie. Jednocześnie ujawniono liczne wyzwania, które obejmują aspekty techniczne, społeczne, etyczne i prawne[103]. Potężna siła transformacyjna i głębokie oddziaływanie na różne dziedziny społeczne AI wywołały szeroką debatę na temat zasad i wartości, które powinny kierować jej rozwojem i wykorzystaniem[104]. Pojawiły się obawy dotyczące wpływu technologii zarówno na jednostkę, w kontekście jej praw podmiotowych, takich jak prywatność, autonomia informacyjna, godność, praw ekonomicznych, w tym odnoszących się do zatrudnienia, praw własności intelektualnej, jak i na całe społeczeństwa, np. w związku z oddziaływaniem na procesy demokratyczne, dostęp do rzetelnej informacji, a w końcu także na środowisko[105]. Źródłem tego oddziaływania jest nie tylko sama specyfika technologii, jej skalowalność technologiczna, zapotrzebowanie na zasoby, również wysokojakościowe, adekwatne dane, ale także sposób jej wykorzystywania. Konsekwencją takiego stanu rzeczy jest także to, że dalszy rozwój wymaga olbrzymich nakładów finansowych, co prowadzi jednocześnie do koncentracji zasobów w rękach kilku globalnych korporacji, utrudniając mniejszym

[102] S. Russell, P. Norvig, *Sztuczna inteligencja*..., s. 50.

[103] W. Stead, *Clinical Implications and Challenges of Artificial Intelligence and Deep Learning*, „JAMA" 2018/320, https://jamanetwork.com/journals/jama/article-abstract/2701665, s. 1107–1108.

[104] A. Jobin, M. Ienca, E. Vayena, *Artificial Intelligence*..., *passim*; E. Awad i in., *The Moral Machine*..., s. 59–63; T. Timan, Z. Mann, *Data Protection*..., s. 165 i n.; J. Fjeld, N. Achten, H. Hilligoss, A.C. Nagy, M. Srikumar, *Principled Artificial*..., *passim*.

[105] M. Brundage i in., *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*, 2018, https://arxiv.org/pdf/1802.07228 (dostęp: 18.12.2024 r.).

podmiotom, m.in. państwom rozwijającym się, pełnoprawne uczestnictwo w wyścigu technologicznym[106]. Niedostateczna jest wreszcie transparentności procesów technologicznych, jak również poziom wiedzy użytkowników, którzy wchodzą w interakcję z technologią. Są to zatem warunki niesprzyjające budowaniu nadszarpniętego zaufania tak do technologii, jak i do podmiotów ją tworzących[107]. Podejście, że „niczego w życiu nie należy się bać, należy to tylko zrozumieć"[108], jest niewystarczające, aby wyeliminować opisane obawy związane z AI. Jest ono natomiast konieczne dla zdefiniowania kierunków niezbędnych interwencji, w tym regulacyjnych.

Organizacje krajowe i międzynarodowe zareagowały na te obawy społeczne, tworząc komitety eksperckie i grupy robocze *ad hoc* ds. sztucznej inteligencji, którym zleciły opracowanie dokumentów strategicznych. Wśród takich grup przykładowo można wymienić powołaną przez Komisję Europejską Grupę ekspertów wysokiego szczebla ds. sztucznej inteligencji (High-Level Expert Group on Artificial Intelligence, HLEG AI)[109] czy Grupę ekspertów ds. sztucznej inteligencji OECD (Working Party on Artificial Intelligence Governance, AIGO)[110], a także Komitet ds. sztucznej inteligencji Rady Europy (Committee on Artificial Intelligence, CAI)[111], Komisję ds. etyki wiedzy naukowej i technicznej UNESCO (World Commission on Ethics of Scientific Knowledge and Technology, COMEST)[112] oraz powołany przez Sekretarza Generalnego ONZ Organ doradczy wysokiego szczebla ds. sztucznej inteligencji (High-Level AI Advisory Body)[113]. Na poziomie krajowym z kolei można

[106] A. Jobin, M. Ienca, E. Vayena, *Artificial Intelligence..., passim*.

[107] D.R. Desai, J.A. Kroll, *Trust But Verify: A Guide to Algorithms & The Law*, „Harvard Journal of Law & Technology" 2017/31(1), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2959472 (dostęp: 18.12.2024 r.).

[108] Autorką tej maksymy jest M. Skłodowska-Curie.

[109] https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai (dostęp: 18.12.2024 r.).

[110] https://oecd.ai/en/network-of-experts (dostęp: 18.12.2024 r.).

[111] https://rm.coe.int/terms-of-reference-of-the-committee-on-artificial-intelligence-cai-/1680ade00f (dostęp: 18.12.2024 r.).

[112] https://www.unesco.org/en/ethics-science-technology/comest (dostęp: 18.12.2024 r.).

[113] https://www.un.org/sites/un2.un.org/files/231025_press-release-aiab.pdf (dostęp: 18.12.2024 r.).

wskazać m.in.: Grupę roboczą ds. sztucznej inteligencji przy Ministrze Cyfryzacji w Polsce, specjalną Komisję ds. sztucznej inteligencji Izby Lordów Zjednoczonego Królestwa (Committee on Artificial Intelligence of the United Kingdom House of Lords), Grupę doradczą ds. sztucznej inteligencji i społeczeństwa obywatelskiego rządu Japonii (Advisory Board on Artificial Intelligence and Human Society)[114] czy Narodową radę nauki i technologii w USA (The National Science and Technology Council), a zwłaszcza w jej ramach Podkomitet ds. uczenia maszynowego i sztucznej inteligencji (Subcommittee on Machine Learning and Artificial Intelligence)[115]. Równolegle powstały oddolne inicjatywy poszczególnych podmiotów i organizacji branżowych, np. Microsoft[116], Google[117] i SAP[118], które opracowały wytyczne i zasady dotyczące AI. Analogiczne wytyczne i opracowania wydały również stowarzyszenia branżowe i organizacje *non-profit*, np. Association of Computing Machinery (ACM)[119], Access Now i Amnesty International[120], Institute of Electrical and Electronics Engineers (IEEE)[121], Future of Life Institute[122], UNI Global Union[123], Global Partnership on AI (GPAI)[124]. Celem tych działań było zdiagnozowanie obszarów wymagających szczególnej uwagi

[114] https://www8.cao.go.jp/cstp/tyousakai/ai/summary/aisociety_en.pdf (dostęp: 18.12.2024 r.).

[115] https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf (dostęp: 18.12.2024 r.).

[116] https://www.microsoft.com/en-us/ai/principles-and-approach#resources (dostęp: 18.12.2024 r.).

[117] https://blog.google/technology/ai/ai-principles/ (dostęp: 18.12.2024 r.).

[118] https://www.sap.com/documents/2018/09/940c6047-1c7d-0010-87a3-c30de2ffd8ff.html (dostęp: 18.12.2024 r.).

[119] https://www.acm.org/code-of-ethics (dostęp: 18.12.2024 r.).

[120] The Toronto Declaration Protecting the right to equality and non-discrimination in machine learning, https://www.torontodeclaration.org (dostęp: 18.12.2024 r.).

[121] Institute of Electrical and Electronics Engineers (IEEE), Ethically Aligned Design, https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf (dostęp: 20.12.2024 r.).

[122] Asilomar AI Principles, https://futureoflife.org/open-letter/ai-principles/ (dostęp: 18.12.2024 r.).

[123] UNI Global Union, 10 Principles for Ethical Artificial Intelligence, https://uniglobalunion.org/report/10-principles-for-ethical-artificial-intelligence/ (dostęp: 18.12.2024 r.).

[124] https://gpai.ai/projects/responsible-ai/ (dostęp: 18.12.2024 r.).

przy rozwoju AI oraz analiza stanu prawnego pod kątem adekwatności i ewentualne sformułowanie zasad adresujących obawy i wyzwania.

2.2. Centralizacja władzy technologicznej

Wspólnym zagadnieniem, stanowiącym oprócz aspektów technologicznych, źródło formułowanych obaw w kontekście AI, jest obserwacja tendencji rynkowych. Odnosi się ona do centralizacji władzy technologicznej, czyli zjawiska, które nie tylko kształtuje rynek technologiczny, ale również redefiniuje relacje między korporacjami, jednostkami i państwami. Sztuczna inteligencja, jako technologia wymagająca ogromnych zasobów danych, infrastruktury obliczeniowej i zaawansowanych algorytmów, skoncentrowała się w rękach kilku globalnych korporacji, określanych mianem Big Tech[125]. Problem ten, jak wskazują m.in. Sh. Zuboff[126] czy W. Hartzog[127], ma głębokie konsekwencje nie tylko dla rynków technologicznych, ale także dla demokracji, prywatności i praw człowieka.

W swoich opracowaniach Sh. Zuboff opisuje jak korporacje technologiczne wykorzystały potencjał dostępnych im danych na temat użytkowników, aby zdominować rynek danych i zbudować swoją pozycję w ekosystemie AI, tworząc przez lata tzw. nadwyżkę behawioralną.

[125] Zob. E. Ponte i in., *Generative AI Raises Competition Concerns*, 29.06.2023, https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/06/generative-ai-raises-competition-concerns (dostęp: 18.12.2024 r.); szerzej F. G'sell, *Regulating under Uncertainty: Governance Options for Generative AI*, https://cyber.fsi.stanford.edu/content/regulating-under-uncertainty-governance-options-generative-ai (dostęp: 18.12.2024 r.), s. 103 i n.; P. Nemitz, M. Pfeffer, *Prinzip Mensch*, Bonn 2020, s. 52–88. Zob. też co do nakładów finansowych Artificial Intelligence Index Report 2024, https://aiindex.stanford.edu/report/ (dostęp: 18.12.2024 r.).

[126] Sh. Zuboff, *Wiek kapitalizmu inwigilacji*, Poznań 2020; Sh. Zuboff, *Surveillance Capitalism or Democracy? The Death Match of Institutional Orders and the Politics of Knowledge in Our Information Civilization*, „Organization Theory" 2022/3(3), https://journals.sagepub.com/doi/10.1177/26317877221129290 (dostęp: 18.12.2024 r.); Sh. Zuboff, *Big other: Surveillance Capitalism and the Prospects of an Information Civilization*, „Journal of Information Technology" 2015/30(1), s. 75.

[127] W. Hartzog, *Two AI Truths*..., s. 612 i n.

Zjawisko to nazwała kapitalizmem inwigilacji, w którym dane użytkowników – określane przez autorkę jako surowce behawioralne – są zbierane, analizowane i przekształcane w predykcyjne produkty, służące do przewidywania oraz wpływania na zachowania jednostek. Niebagatelną rolę odgrywa tutaj AI, która przetwarza olbrzymie ilości danych w celu generowania prognoz i modeli zachowań, często bez zgody użytkowników i w sposób sprzeczny z zasadami ochrony danych[128]. Jak twierdzi Sh. Zuboff, dominacja Big Tech, takich jak Google, Meta czy Amazon, pozwoliła tym podmiotom przejąć kontrolę nad kluczowymi elementami infrastruktury technologicznej, w tym nad danymi, algorytmami i chmurami obliczeniowymi. W rezultacie korporacje te stały się nie tylko liderami innowacji, ale także strażnikami, którzy określają, kto i w jaki sposób może korzystać z nowoczesnych technologii. Autorka podkreśla, że ta asymetria władzy podważa fundamentalne zasady równości i demokracji, ponieważ decyzje dotyczące wykorzystania danych i algorytmów podejmowane są przez podmioty prywatne, kierujące się przede wszystkim własnym interesem ekonomicznym[129].

Analizę Sh. Zuboff rozszerza W. Hartzog, wskazując na główne niebezpieczeństwa związane z koncentracją władzy technologicznej[130]. Krytycznie odnosi się on do powszechnej narracji o neutralności i obiektywizmie AI, wskazując, że technologie te są projektowane i wdrażane w sposób zgodny z interesami korporacji, a nie koniecznie z potrzebami społeczeństwa. Zjawisko to prowadzi do trzech istotnych problemów, tj. według autora – „dwóch prawd i jednego kłamstwa". Po pierwsze trzeba wskazać na tendencje do maksymalizacji zysków przez intensyfikację gromadzenia danych oraz wykorzystywanie zasobów ludzkich i środowiskowych. W efekcie użytkownicy tracą kontrolę nad swoimi danymi, a ich prywatność jest systematycznie naruszana. Po wtóre wykorzystywane są tendencje społeczne do stopniowej akceptacji inwazyjnych praktyk technologicznych, np. masowego monitorowana czy

[128] Sh. Zuboff, *Surveillance...*, *passim*.

[129] Sh. Zuboff, *Surveillance...*, *passim*; zob. również P. Pałka, *Ciało obce: zasady RODO a gospodarka rynkowa*, „internetowy Kwartalnik Antymonopolowy i Regulacyjny" 2022/11(5), https://press.wz.uw.edu.pl/ikar/vol11/iss5/4/ (dostęp: 18.12.2024 r.), s. 62–84.

[130] W. Hartzog, *Two AI Truths...*, s. 612 i n.

algorytmicznego profilowana, co prowadzi do osłabienia norm dotyczących prywatności i etyki. Po trzecie, co stanowi owo kłamstwo, rozwój technologiczny odbywa się w celu maksymalizacji korzyści społecznych. Autor podkreśla, że chociaż prawdą jest, iż ludzie prawdopodobnie odniosą pewne korzyści, to jednak nie to jest zasadniczym celem, a jedynie pretekstem do ekspansji rynkowej[131]. Zjawiska te nawiązują do koncepcji cyfrowego autorytaryzmu (*digital authoritarianism*), przeniesionej z poziomu politycznego na poziom technologiczny i ekonomiczny, w którym technologie AI są wykorzystywane do kontrolowania i manipulowania zachowaniami jednostek oraz całych społeczności[132]. W połączeniu z brakiem przejrzystości i odpowiedzialności korporacji technologicznych, centralizacja władzy technologicznej staje się zagrożeniem nie tylko dla jednostek, ale także dla demokracji i zasad rządów prawa. Dominacja Big Tech w obszarze zarządzania danymi i algorytmami pozwala korporacjom wpływać na dyskurs publiczny, moderować treści w mediach społecznościowych oraz manipulować algorytmami rekomendacji. Praktyki te nie tylko kształtują opinie publiczne, ale także mogą być wykorzystywane do promowania określonych interesów politycznych czy ekonomicznych, co podważa zasady pluralizmu i równości w demokratycznym procesie decyzyjnym. Przykładem tego rodzaju wpływu jest wykorzystywanie algorytmów do tworzenia tzw. baniek informacyjnych (*information bubbles*), które izolują użytkowników od treści sprzecznych z ich poglądami, co prowadzi do polaryzacji społeczeństwa[133]. Ponadto brak przejrzystości w moderacji treści oraz działaniu algorytmów uniemożliwia obywatelom i instytucjom demokratycznym skuteczny nadzór nad działaniami korporacji technologicznych.

[131] W. Hartzog, *Two AI Truths*..., s. 612 i n.

[132] J.S. Pearson, *Defining Digital Authoritarianism*, „Philosophy & Technology" 2024/37, https://www.researchgate.net/publication/381324260_Defining_Digital_Authoritarianism (dostęp: 30.12.2024 r.).

[133] O. Kuzmenko, A. Cyburt, H. Yarovenko, V. Yersh, Y. Humenna, *Modeling of 'information bubbles' in the global information space*, „Journal of International Studies" 2021/14(4), https://www.researchgate.net/publication/357729835_Modeling_of_information_bubbles_in_the_global_information_space (dostęp: 30.12.2024 r.), s. 270–285.

Z powyższych względów już na płaszczyźnie wyzwań rynkowych związanych z centralizacją władzy technologicznej konieczne jest wprowadzenie lub skuteczne egzekwowanie odpowiednich regulacji, które mają ograniczyć dominację rynkową niektórych podmiotów i zapewnić rozwój AI w zgodzie z zasadami demokratycznymi i interesem publicznym[134].

Opisany problem ma charakter systemowy, wymagający kompleksowego podejścia regulacyjnego, które jest konieczne dla zatrzymania pogłębiania się nierówności, osłabiania procesów demokratycznych i naruszania podstawowych praw człowieka[135]. Kierunek ten, na co zwraca uwagę A. Bradford, mimo podnoszonych głosów, że regulacje hamują innowacje, jest kluczowy dla zrównoważonego rozwoju[136]. Może on także pozytywnie oddziaływać na innowacje, zwłaszcza w obszarach takich jak ochrona prywatności i bezpieczeństwo danych, stymulując rozwój bardziej zrównoważonych modeli biznesowych[137]. Zgadzam się z powołaną autorką, że państwa nie muszą wybierać między regulacją a innowacją[138]. Istotne jest jednak równoległe wdrażanie reform instytucjonalnych, które wspierają innowacyjność, zapewniając lepszą integrację rynku, poprawę dostępu do finansowania i rozwój kultury przedsiębiorczości.

2.3. Wyzwania etyczne

W dużej mierze, w odpowiedzi na potencjał technologii i jej negatywnego oddziaływania zarówno indywidualnego, jak i społecznego, które

[134] Można przykładowo sięgnąć po mechanizmy prawa antymonopolowego.

[135] J. Schuett, *A Legal Definition*..., *passim*.

[136] A. Bradford, *The False Choice Between Digital Regulation and Innovation*, „Northwestern University Law Review" 2024/119(2), https://scholarlycommons.law.northwestern.edu/cgi/viewcontent.cgi?article=1578&context=nulr (dostęp: 30.12.2024 r.), s. 377 i n.

[137] Na temat relacji innowacji i obaw przed przeregulowaniem rynku zob. również M. Draghi, *The future of European competitiveness – A competitiveness strategy for Europe*, https://commission.europa.eu/topics/strengthening-european-competitiveness/eu-competitiveness-looking-ahead_en (dostęp: 30.12.2024 r.).

[138] A. Bradford, *The False Choice*..., s. 377 i n.