OBRAZY GENEROWANE Z WYKORZYSTANIEM SZTUCZNEJ INTELIGENCJI

Status prawnoautorski

Aleksandra Bar

Wolters Kluwer

OBRAZY GENEROWANE Z WYKORZYSTANIEM SZTUCZNEJ INTELIGENCJI

Status prawnoautorski

Aleksandra Bar

Wolters Kluwer

Warszawa 2025

Stan prawny na 1 stycznia 2025 r.

Recenzentka
Dr hab. Ewa Laskowska-Litak

Wydawczyni
Monika Pawłowska

Redaktorka prowadząca
Kinga Zając

Opracowanie redakcyjne
Trzy kropki Joanna Maź

Projekt okładek serii
Wojtek Janikowski

ISBN 978-83-8390-406-1 *ISBN PDF-a: 978-83-8390-450-4*

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 9

Przedmowa 13

Wprowadzenie 15

Rozdział 1
Sztuczna inteligencja 21
1.1. Wstęp 21
1.2. Czym jest sztuczna inteligencja? 21
1.3. Dzieje sztucznej inteligencji 24
1.3.1. Pierwsze sukcesy i porażki 24
1.3.2. Droga ku głębokiemu uczeniu 28
1.3.2.1. Sieci neuronowe 28
1.3.2.2. Świt głębokiego uczenia 30
1.3.2.3. Uczenie nadzorowane i nienadzorowane 34
1.4. Obecny stan sztucznej inteligencji 35
1.4.1. Wiele twarzy głębokiego uczenia 35
1.4.2. Źródła sukcesu 36
1.4.3. Ogólna i wąska sztuczna inteligencja 38
1.5. Filozofia sztucznej inteligencji 39
1.5.1. Wielkie nadzieje i obawy 39
1.5.2. Czy komputery potrafią myśleć? Test Turinga i „chiński pokój” 40
1.6. Definiowanie sztucznej inteligencji 43
1.6.1. Definiowanie sztucznej inteligencji dla celów regulacyjnych 43
1.6.2. Definiowanie sztucznej inteligencji dla celów analizy prawnej 45
1.6.3. Sztuczna inteligencja w rozumieniu monografii – wyjaśnienie pojęcia 50
1.7. Automatyzacja twórczości 55
1.7.1. Wstęp 55
1.7.2. Czy sztuczna inteligencja może być twórcza? Twórczość AI w ujęciu pozaprawnym 58

1.7.3. Wybrane osiągnięcia w dziedzinie sztuk plastycznych 62
1.7.3.1. AARON 63
1.7.3.2. Artystyczna stylizacja obrazów i problem „transferu stylu” 64
1.7.3.3. *Nowy Rembrandt* 71
1.7.3.4. Generatywne sieci antagonistyczne 74
1.7.3.5. Modele dyfuzyjne jako nowa jakość w dziedzinie tłumaczenia tekstu na obraz 83
1.7.4. Podsumowanie 86

Rozdział 2
Prawo autorskie 91
2.1. Przedmiot prawa autorskiego 91
2.1.1. Wstęp 91
2.1.2. Utwór w świetle przepisów ustawy o prawie autorskim i prawach pokrewnych 93
2.1.2.1. Przedmiotowe przesłanki ochrony prawnoautorskiej 94
2.1.2.2. Brak ochrony pomysłów 121
2.1.2.3. Okoliczności irrelewantne 132
2.1.2.4. O kryzysie pojęcia utworu 141
2.1.3. Pojęcie utworu w prawie Unii Europejskiej 144
2.2. Podmiot prawa autorskiego 149
2.2.1. Wstęp 149
2.2.2. Twórca 150
2.2.2.1. Uwagi ogólne 150
2.2.2.2. Zasada ludzkiego autorstwa 152
2.2.3. Współtwórca 154
2.2.3.1. Uwagi ogólne 154
2.2.3.2. Wniesienie wkładu twórczego 155
2.2.3.3. Porozumienie współtwórców 162
2.2.4. Efekty działania sił przyrody i zwierząt a zasada ludzkiego autorstwa 163

Rozdział 3
Dzieła komputerowe w świetle prawa autorskiego 169
3.1. Dzieła komputerowe 169
3.1.1. Pojęcie 169
3.1.2. Podział 172
3.1.2.1. Dyskusja nad rolą technologii komputerowej w procesie kreacyjnym – uwagi terminologiczne 172
3.1.2.2. Dzieła wspomagane komputerowo i dzieła generowane komputerowo 183
3.2. Dzieła wspomagane komputerowo 186
3.2.1. Podmioty potencjalnie autorsko uprawnione 187

3.2.1.1. Faza rozwojowa ... 188
3.2.1.2. Faza eksploatacyjna ... 194
3.2.1.3. Podsumowanie ... 196
3.2.2. Dzieła wspomagane komputerowo odzwierciedlające twórczy wkład użytkownika programu ... 196
3.2.2.1. Programy-narzędzia ... 200
3.2.2.2. Programy-kreatory ... 222
3.2.2.3. Programy-generatory ... 227
3.2.2.4. Podsumowanie ... 256
3.2.3. Dzieła wspomagane komputerowo odzwierciedlające twórczy wkład twórcy programu ... 258
3.2.3.1. Systemy autonomiczne ... 259
3.2.3.2. Systemy quasi-autonomiczne ... 265
3.2.3.3. Podsumowanie ... 283
3.3. Dzieła generowane komputerowo ... 284
3.4. Dzieła wspomagane komputerowo a dzieła generowane komputerowo – trudności w procesie stosowania prawa ... 285

Rozdział 4
Dzieła generowane komputerowo w polskim porządku prawnym ... 289
4.1. Zamiast wstępu. Dzieła generowane komputerowo jako nieczyste dobra publiczne – perspektywa ekonomiczna ... 289
4.2. Uwagi *de lege lata* ... 292
4.2.1. Dzieła generowane komputerowo jako dobra cyfrowe – problem zabezpieczeń technicznych ... 292
4.2.2. Dzieła generowane komputerowo jako niematerialne dobra prawne ... 295
4.2.2.1. Dzieła generowane komputerowo jako przedmioty niematerialne ... 295
4.2.2.2. Dzieła generowane komputerowo jako dobra prawne ... 296
4.2.3. Dzieła generowane komputerowo jako przedmiot praw na dobrach niematerialnych ... 297
4.2.4. Dzieła generowane komputerowo jako element domeny publicznej a odpowiedzialność za działania podejmowane względem tych dzieł ... 300
4.2.4.1. Odpowiedzialność deliktowa ... 303
4.2.4.2. Odpowiedzialność kontraktowa ... 340
4.2.5. Podsumowanie ... 356
4.3. Uwagi *de lege ferenda* ... 357

Zakończenie ... 381

Bibliografia ... 385

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)
k.c.	–	ustawa z 23.04.1964 r. – Kodeks cywilny (Dz.U. z 2024 r. poz. 1061 ze zm.)
Konwencja berneńska	–	Akt paryski konwencji berneńskiej o ochronie dzieł literackich i artystycznych, sporządzony w Paryżu 24.07.1971 r. (Dz.U. z 1990 r. Nr 82, poz. 474)
pr. aut. / ustawa autorska	–	ustawa z 4.02.1994 r. o prawie autorskim i prawach pokrewnych (Dz.U. z 2025 r. poz. 24 ze zm.)
p.w.p.	–	ustawa z 30.06.2000 r. – Prawo własności przemysłowej (Dz.U. z 2023 r. poz. 1170)
u.z.n.k.	–	ustawa z 16.04.1993 r. o zwalczaniu nieuczciwej konkurencji (Dz.U. z 2022 r. poz. 1233)

Publikatory

AUS	–	Acta Iuris Stetinensis
Dz.U.	–	Dziennik Ustaw
Dz.Urz. UE C	–	Dziennik Urzędowy Unii Europejskiej Seria C
Dz.Urz. UE L	–	Dziennik Urzędowy Unii Europejskiej Seria L
EPS	–	Europejski Przegląd Sądowy
IKAR	–	Internetowy Kwartalnik Antymonopolowy i Regulacyjny
IPQ	–	Intellectual Property Quarterly
KPP	–	Kwartalnik Prawa Prywatnego
MoP	–	Monitor Prawniczy
NP	–	Nowe Prawo
ONSAiWSA	–	Orzecznictwo Naczelnego Sądu Administracyjnego i Wojewódzkich Sądów Administracyjnych
OSA	–	Orzecznictwo Sądów Apelacyjnych
OSN	–	Orzecznictwo Sądu Najwyższego
OSNC	–	Orzecznictwo Sądu Najwyższego. Izba Cywilna
OSNCP	–	Orzecznictwo Sądu Najwyższego Izby Cywilnej, Pracy i Ubezpieczeń Społecznych

OSN(K)	–	Orzeczenia Sądu Najwyższego (Izba Karna)
OSNP	–	Orzecznictwo Sądu Najwyższego. Izba Pracy, Ubezpieczeń Społecznych i Spraw Publicznych
OSP	–	Orzecznictwo Sądów Polskich
OTK	–	Orzecznictwo Trybunału Konstytucyjnego
PiP	–	Państwo i Prawo
PME	–	Prawo Mediów Elektronicznych
PNT	–	Prawo Nowych Technologii – dodatek specjalny do Monitora Prawniczego
PPH	–	Przegląd Prawa Handlowego
REStat	–	The Review of Economics and Statistics
ZNUJ PPWI	–	Zeszyty Naukowe Uniwersytetu Jagiellońskiego. Prace z Prawa Własności Intelektualnej
ZNUJ PWiOWI	–	Zeszyty Naukowe Uniwersytetu Jagiellońskiego. Prace z Wynalazczości i Ochrony Własności Intelektualnej

Organy i instytucje

AIPPI UK	–	brytyjska Grupa Międzynarodowego Stowarzyszenia Ochrony Własności Intelektualnej (ang. International Association for the Protection of Intellectual Property United Kingdom)
CONTU	–	Komisja do spraw Nowych Zastosowań Technologicznych Dzieł Chronionych Prawami Autorskimi (ang. National Commission on New Technological Uses of Copyrighted Work (US))
DARPA	–	Agencja Zaawansowanych Projektów Badawczych w Obszarze Obronności (ang. Defence Advanced Research Projects Agency)
NSA	–	Naczelny Sąd Administracyjny
OTA	–	Urząd do spraw Oceny Technologii (ang. Office of Technology Assessment (US))
SA	–	sąd apelacyjny
SN	–	Sąd Najwyższy
TK	–	Trybunał Konstytucyjny
TS	–	Trybunał Sprawiedliwości
WIPO	–	Światowa Organizacja Własności Intelektualnej (ang. World Intellectual Property Organization)
WSA	–	wojewódzki sąd administracyjny

Inne

AI	–	sztuczna inteligencja (ang. *Artificial Intelligence*)
ALT	–	tekst alternatywny (ang. *Alternative Text*)
ANN	–	sztuczne sieci neuronowe (ang. *Artificial Neural Network*)
AttnGAN	–	*Attentional Generative Adversarial Network*
BigGAN	–	*Big Generative Adversarial Network*
CAD	–	system komputerowo wspomaganego projektowania (ang. *Computer Aided Design*)
CAN	–	kreatywne sieci antagonistyczne (ang. *Creative Adversarial Network*)
cGAN	–	warunkowa architektura GAN (ang. *Conditional Generative Adversarial Network*)
CLIP	–	*Contrastive Language-Image Pre-training*

CUDA	–	*Compute Unified Device Architecture*
DCGAN	–	głęboka splotowa sieć GAN (ang. *Deep Convolutional Generative Adversarial Network*)
DF-GAN	–	*Deep Fusion Generative Adversarial Network*
DM-GAN	–	*Dynamic Memory Generative Adversarial Network*
DNN	–	głębokie sieci neuronowe (ang. *Deep Neural Network*)
EAP	–	ekonomiczna analiza prawa
GAN	–	generatywne sieci antagonistyczne (ang. *Generative Adversarial Network*)
GPU	–	układy graficzne (ang. *Graphics Processing Unit*)
NFT	–	niewymienny token (ang. *Non-fungible Token*)
PGAN	–	progresywny wzrost GAN (ang. *Progressive Growing of Generative Adversarial Network*)
RAM	–	*Random Access Memory*
RefineGAN	–	antagonistyczne sieci przeciwstawne (ang. *Refined Generative Adversarial Networks*)
SaaS	–	model „oprogramowanie jako usługa" (ang. *Software as a Service*)
StackGAN	–	*Stacked Generative Adversarial Network*
StyleGAN	–	*Style-Based Generative Adversarial Network*
VQGAN	–	*Vector Quantized Generative Adversarial Network*

PRZEDMOWA

Niniejsza monografia stanowi zmodyfikowaną wersję rozprawy doktorskiej obronionej przez autorkę na Wydziale Prawa, Administracji i Ekonomii Uniwersytetu Wrocławskiego w marcu 2024 r.

Za życzliwość oraz wsparcie udzielone w czasie przygotowywania rozprawy serdecznie dziękuję promotorowi – Panu prof. dr. hab. Piotrowi Machnikowskiemu. Podziękowania kieruję także do recenzentów rozprawy – Pana prof. dr. hab. Ryszarda Markiewicza, Pani prof. dr hab. Katarzyny Grzybczyk oraz Pana dr. hab. Wojciecha Machały, których cenne uwagi pomogły w zredagowaniu ostatecznej wersji opracowania. Serdecznie dziękuję też Stowarzyszeniu Autorów i Wydawców Copyright Polska za wsparcie stypendialne udzielone na etapie przygotowywania rozprawy doktorskiej.

Dziękuję wszystkim osobom zaangażowanym w proces wydawniczy tej książki, w szczególności Pani Monice Pawłowskiej oraz Pani Kindze Zając. Za trud włożony w przygotowanie recenzji wydawniczej serdecznie dziękuję Pani dr hab. Ewie Laskowskiej-Litak.

Moim bliskim – rodzinie i przyjaciołom – składam wyrazy wdzięczności za zrozumienie i cierpliwość. Mężowi, który był dla mnie ogromnym wsparciem w każdej trudnej chwili, dziękuję szczególnie. Miłoszu, bez Ciebie nie powstałaby ta praca. Tobie tę książkę dedykuję.

WPROWADZENIE

Twórczość w erze cyfrowej przybiera nowe wcielenia, które nieodłącznie związane są z rozwojem technologii. Eksploracja kreatywnego potencjału obliczeniowych technik komputerowych doprowadziła do wykształcenia się zjawiska określanego w niniejszej monografii mianem twórczości zautomatyzowanej. Pojęciem tym posługuję się w kontekstach pozaprawnych dla opisu fenomenu wyrosłego na styku dwóch sfer – technologii i kultury. W przyjętym przeze mnie rozumieniu twórczość zautomatyzowana to każde przedsięwzięcie, w ramach którego program komputerowy, działając z pewnym zakresem niezależności od człowieka, generuje obiekt, który rozpoznajemy – nie będąc świadomymi komputerowej proweniencji ocenianego rezultatu – jako nowy i wartościowy wytwór myśli ludzkiej.

Tak rozumiana twórczość zautomatyzowana od wielu lat wzbudza – w wymiarze globalnym – zainteresowanie przedstawicieli nauki prawa autorskiego. Zgodnie z poglądem tradycyjnym, wspólnym dla większości współczesnych systemów prawa autorskiego (a przynajmniej właściwym dla systemów państw-sygnatariuszy Konwencji berneńskiej), aby określony wytwór podlegał ochronie autorskiej, niezbędne jest ustalenie, że jego twórcą jest człowiek. Stwierdzenie braku twórcy-człowieka przesądza o pozostawieniu danego obiektu poza zakresem ochrony prawa autorskiego. Pojawienie się pierwszych komputerowych systemów informacyjnych w drugiej połowie ubiegłego wieku, a wraz z nimi programów realizujących zadania z zakresu przetwarzania i generowania treści w ramach procesu, nad którymi żaden człowiek nie sprawuje bezpośredniej kontroli, zrodziło konieczność zmierzenia się z pytaniem o status prawnoautorski ich rezultatów. Amerykańscy i europejscy badacze prawa autorskiego końca XX w. odrzucali myśl, jakoby znane w tamtym czasie systemy komputerowe były rzeczywistym wyzwaniem (zarówno z punktu widzenia przepisów prawa stanowionego, jak i aksjologii regulacji ochrony twórczości), wieszcząc jednak rychłe pojawienie się programów, których działanie ostatecznie miałoby podważyć paradygmat twórcy-człowieka.

Prognozowana w tamtych czasach rewolucja nie nadeszła może tak szybko, jak wówczas zakładano, ale wydaje się, że oto stajemy dziś w obliczu zjawiska, przed którym nas niegdyś „ostrzegano". Ostatnie osiągnięcia w dziedzinie uczenia maszynowego, która stanowi najprężniej rozwijający się obszar badań nad tzw. sztuczną inteligencją (ang. *Artificial Intelligence*, AI), stanowią siłę napędową szerokiego wachlarza prze-

łomowych rozwiązań i tworzą podstawę wielu zastosowań w dziedzinie twórczości zautomatyzowanej. Jednocześnie skala komputeryzacji i internetyzacji współczesnych społeczeństw sprawia, że tego rodzaju osiągnięcia rozprzestrzeniają się w bezprecedensowym tempie.

Powyższe skłania do uznania, że samouczące systemy AI wykazujące zdolność do generowania twórczości zautomatyzowanej, których rezultaty rozpalają dziś dyskurs prawa autorskiego, nie są z punktu widzenia prawa problemem nowym[1]. Natomiast trudno zaprzeczyć, że w związku z upowszechnieniem się tego rodzaju technologii pytanie o status prawnoautorski twórczości zautomatyzowanej zyskało zupełnie nowy wymiar. Każdego dnia, dzięki wykorzystaniu programów korzystających z technologii uczenia maszynowego w przestrzeni wirtualnej, generowane są miliony nowych obrazów, które nie dają się odróżnić od tych tworzonych przez ludzi z wykorzystaniem tradycyjnych i cyfrowych narzędzi artystycznej kreacji. Stajemy zatem przed problemem „zalewu" przestrzeni twórczej dziełami, których status prawnoautorski jest co najmniej wątpliwy z uwagi na trudności ze wskazaniem twórcy-człowieka. Jednocześnie można przypuszczać, że w najbliższej przyszłości samouczące systemy AI, które posiadają zdolność do generowania twórczości zautomatyzowanej i które już dziś są powszechnie dostępne, będą szeroko wykorzystywane w praktyce rynkowej (chociażby w branży rozrywkowej czy reklamowej). Uzasadnione jest zatem oczekiwanie, że w przypadku stwierdzenia braku ochrony autorskiej podmioty, które w związku z zaistnieniem dzieł wytwarzanych przez te systemy ponoszą określone nakłady, czynić będą starania o zwrot poczynionych inwestycji finansowych. Przejawem tych starań może być dążenie do objęcia twórczości zautomatyzowanej ochroną prawną, w tym – w przypadku stwierdzenia braku adekwatnych narzędzi ochrony na gruncie obowiązującego prawa – podejmowanie przez zainteresowane grupy interesów działań zmierzających do stworzenia nowej kategorii praw wyłącznych o bezwzględnym charakterze. Stwierdzenie braku ochrony na gruncie prawa autorskiego przy jednoczesnym braku jakiejkolwiek interwencji ustawodawczej może rodzić natomiast ryzyko wystąpienia zjawiska polegającego na przypisywaniu tego rodzaju dzieł własnej działalności twórczej przez osoby fizyczne, a w konsekwencji zagrażać twórczości ludzkiej.

Z uwagi na to, że wytwory będące przejawami twórczości zautomatyzowanej wykazują często takie właściwości, które gdyby zostały nadane przez człowieka, uzasadniałyby przyznanie im statusu utworów w rozumieniu ustawy o prawie autorskim i prawach pokrewnych, kluczowe jest rozstrzygnięcie w przedmiocie prawnoautorskiego statusu tego rodzaju obiektów. Skoro stwierdzenie braku twórcy-człowieka przesądza o pozostawieniu danego obiektu poza zakresem ochrony prawa autorskiego, w odniesieniu do wytworów będących przejawami twórczości zautomatyzowanej konieczne jest ustalenie, czy w danym obiekcie zaznacza się twórczy wkład człowieka. Owo „poszukiwanie"

[1] Choć, jak zostanie to wykazane w opracowaniu, systemy uczenia maszynowego stanowią pod wieloma względami „nową jakość" na tle programów, które nie korzystają z tego rodzaju technologii.

osoby twórczo odpowiedzialnej za zaistnienie danego rezultatu wyznacza strukturę prowadzonych w opracowaniu rozważań.

Nie wyprzedzając zanadto dalszego wywodu, wypada wyjaśnić, że w monografii przyjęto szerokie rozumienie sztucznej inteligencji. Pojęcia tego nie odnoszę wyłącznie do systemów korzystających z metod uczenia maszynowego, ale obejmuję nim wszystkie rozwiązania informatyczne, których nośnikiem jest taki program komputerowy, który działając z pewnym zakresem niezależności od człowieka, generuje określone wyniki, przy czym niezależność programu ujawnia się co najmniej na poziomie rezultatów jego działania. Występuje zatem zbieżność pomiędzy przyjętą definicją sztucznej inteligencji a pojęciem twórczości zautomatyzowanej, uprawniająca wniosek, że wszystkie systemy zdolne do generowania twórczości zautomatyzowanej są w istocie systemami AI.

Przedmiotem badań uczyniłam jedynie takie wytwory AI, które mają postać wizualną i dwuwymiarową (są obrazami cyfrowymi). Co do zasady, ustawodawca nie różnicuje zasad ochrony utworu w zależności od tego, do jakiej kategorii twórczości utwór należy. Dokonywana w procesie stosowania prawa ocena spełnienia przesłanek ochrony autorskiej dzieł należących do różnych kategorii twórczości musi jednak uwzględniać specyfikę danej dziedziny działalności kreacyjnej. To, w jakich elementach może się przejawiać twórczy i indywidualny charakter, jest bowiem ściśle związane z tym, do jakiej dziedziny twórczości należy oceniany obiekt. Potrzeba utrzymania monografii w racjonalnych rozmiarach przesądziła więc o ograniczeniu pola badawczego do jednej, wybranej dziedziny twórczości. Decyzja o ograniczeniu badań do takich tylko wytworów AI, które mają postać wizualną i dwuwymiarową – a więc wytworów należących do kategorii form płaszczyznowych twórczości plastycznej – motywowana była kompetencjami autorki oraz faktem, że rezultaty osiągane przez AI w tej dziedzinie działalności kreacyjnej są najbardziej imponujące i zostały najszerzej opisane w literaturze[2].

Co zatem istotne, twierdzenia formułowane w niniejszej pracy nie są odnoszone do całej – bądź co bądź niezwykle obszernej i wewnętrznie zróżnicowanej – kategorii obiektów tworzonych z wykorzystaniem AI. Choć konkluzje wynikające z rozważań zawartych w monografii mogą nadawać się do rozciągnięcia na wytwory AI należące do innych dziedzin (np. muzyczne czy piśmiennicze), wszelkie uogólnienia muszą być dokonywane ostrożnie i powinny być poprzedzone refleksją w zakresie specyficznych dla tych dziedzin przejawów twórczości oraz analizą reprezentatywnych osiągnięć AI w tych obszarach. Podsumowując tę część wywodu, wypada zaznaczyć, że choć zasadniczym impulsem do podjęcia problematyki niniejszego opracowania były sygnalizowane w piśmiennictwie wątpliwości w zakresie prawnoautorskiego statusu twórczości zautomatyzowanej w ogóle, to formułowane w publikacji twierdzenia nie obejmują, a przynajmniej nie wprost, wytworów AI innych niż obrazy cyfrowe.

[2] Na moment podjęcia prac nad niniejszym opracowaniem (2019 r.).

Refleksja nad zjawiskiem twórczości zautomatyzowanej czyniona przez pryzmat roli współczesnych technik komputerowych w wytwarzaniu dwuwymiarowych dzieł plastycznych pozwoliła już na wstępie na sformułowanie ogólnej tezy, że nie wszystkie obrazy cyfrowe powstałe dzięki wykorzystaniu systemów AI charakteryzują się takim samym statusem prawnoautorskim z uwagi na fakt, że różna jest funkcja, jaką program sztucznej inteligencji może pełnić w procesie kreacyjnym. Teza ta jest wynikiem obserwacji, że w praktyce występuje szerokie spektrum scenariuszy współdzielonej odpowiedzialności człowieka i systemu AI za przebieg procesu generatywnego i jego rezultat. Powyższa obserwacja, wspierana dostrzeżeniem licznych wątpliwości związanych z definiowaniem sztucznej inteligencji, przesądziła ostatecznie o przyjęciu szerokiego kontekstu badania: problematykę prawnoautorskiego statusu twórczości zautomatyzowanej podejmuję w niniejszym opracowaniu na tle ogólniejszego problemu wpływu wykorzystania technik komputerowych w procesie kreacyjnym na możliwość stwierdzenia występowania twórczego wkładu człowieka w dziele. Osadzenie rozważań dotyczących statusu twórczości zautomatyzowanej w szerszym, technologicznym i prawnym kontekście wytworów tworzonych z wykorzystaniem technologii komputerowej w ogóle (tzw. dzieł komputerowych) umożliwiło poszukiwanie odniesień i analogii w dotychczasowych wypowiedziach doktryny, formułowanych na tle znanych od lat rozwiązań, co z założenia miało pozwolić na uniknięcie tendencji do przeceniania rzeczywistej roli systemów AI w procesie generatywnym, która to tendencja ściśle wiąże się z ogólną inklinacją do przypisywania im większej sprawczości, aniżeli jest w rzeczywistości.

Rozważaniom przedstawionym w monografii przyświecały dwa cele badawcze. Zasadniczym celem było stwierdzenie, czy na gruncie przepisów polskiego prawa, w świetle dorobku nauki prawa autorskiego i aktualnych wypowiedzi orzecznictwa, wizualne i dwuwymiarowe wytwory AI mogą podlegać ochronie jako utwory w rozumieniu ustawy o prawie autorskim i prawach pokrewnych, czy też funkcja programów AI w procesie kreacyjnym wyłącza możliwość zaznaczenia się w tych obrazach twórczego wkładu człowieka. Ubocznym celem rozważań jest natomiast ustalenie – w przypadku stwierdzenia, że wizualne i dwuwymiarowe wytwory AI nie podlegają *de lege lata* ochronie autorskiej – czy należy chronić te ostatnie prawami wyłącznymi, w szczególności prawami własności intelektualnej, i sformułowanie w tym zakresie postulatów *de lege ferenda*.

Zagadnienie prawnoautorskiego statusu treści powstających z wykorzystaniem AI nie wyczerpuje bynajmniej szerokiego wachlarza problemów, z jakimi przychodzi zmierzyć się nauce prawa własności intelektualnej w obliczu postępującego rozwoju technologii sztucznej inteligencji. Trzeba zatem w tym miejscu wskazać obszary, które choć rozpalają w ostatnim czasie dyskurs naukowy i rodzą wiele wątpliwości w praktyce, nie zostały objęte analizą. Po pierwsze, w opracowaniu nie podejmuję problematyki granic swobody eksploatacji przedmiotów praw wyłącznych na potrzeby trenowania algorytmów. Po drugie, nie odnoszę się do pytania o konieczność sprawiedliwego wynagradzania twórców z tytułu eksploatacji ich utworów oraz innych przedmiotów praw wyłącznych dla potrzeb tworzenia modeli generatywnych. Po trzecie, nie podejmuję rozważań w zakresie

przyszłości prawa autorskiego w ogólności (nie rozważam kwestii zasadności utrzymania go w obecnym kształcie, w szczególności gdy idzie o zakres ochrony utworów czy stopień twórczości konieczny dla stwierdzenia istnienia przedmiotu ochrony).

Monografia składa się z wprowadzenia, czterech rozdziałów tematycznych oraz zakończenia.

Celem rozdziału pierwszego jest omówienie zjawiska sztucznej inteligencji według technicznych i filozoficznych aspektów jej funkcjonowania oraz refleksja nad historią „inteligentnych" maszyn. W rozdziale pierwszym podjęłam się zdefiniowania pojęcia sztucznej inteligencji oraz przedstawiłam obecny stan rozwoju AI, poświęcając szczególne miejsce w tym zakresie dotychczasowym osiągnięciom w dziedzinie twórczości plastycznej.

W rozdziale drugim, w zakresie niezbędnym z punktu widzenia celów opracowania, dokonano dogmatycznej analizy przedmiotowych przesłanek ochrony prawnoautorskiej w zakresie pozwalającym na ustalenie, jakie cechy powinien, w myśl generalnej definicji utworu rekonstruowanej na podstawie przepisów ustawy o prawie autorskim i prawach pokrewnych oraz w świetle reprezentatywnych poglądów doktryny i orzecznictwa w zakresie wykładni ustawowych kryteriów ochrony, posiadać obiekt, aby w świetle polskiego prawa mógł być uznany za przedmiot prawa autorskiego. W tym rozdziale podjęłam się też ustalenia przesłanek, które można uznać za decydujące i konieczne dla nabycia statusu twórcy lub współtwórcy utworu. Podstawowym celem rozważań w tym przedmiocie było omówienie pojęcia wkładu twórczego oraz rozważenie, w świetle wypowiedzi przedstawicieli polskiej nauki prawa autorskiego i judykatury, jakie rodzaje ludzkiej aktywności, które – chociaż są niekiedy niezbędne dla zaistnienia utworu – nie predestynują do miana jego (współ)twórcy. Odrębne miejsce w tej części monografii poświęciłam omówieniu zasady ludzkiego autorstwa oraz zagadnieniu prawnoautorskiej ochrony efektów działania sił przyrody i zwierząt, które na tle rozważań w zakresie prawnoautorskiego statusu wytworów AI ma szczególny charakter[3].

Rozważania rozdziału trzeciego rozpoczęłam od określenia treści nazwy „dzieło komputerowe", to jest ustalenia, jakie są warunki konieczne i wystarczające do tego, by uznać obiekt za desygnat tej nazwy. Kolejno podjęłam próbę systematyzacji terminologii wykształconej na gruncie badań nad problematyką prawnoautorskiego statusu dzieł powstających z wykorzystaniem technologii komputerowej. Następnie dokonałam klasyfikacji dzieł komputerowych, opierając się na kryterium prawnoautorskiego statusu desygnatów tego pojęcia, przeprowadzając dychotomiczny podział dzieł komputero-

[3] Rozważania dotyczące problematyki AI często prowadzone są przez porównanie systemów AI do zwierząt. Za cechę wspólną systemów AI i zwierząt uważa się ich niezależność od człowieka. W kontekście ocen prawnoautorskich porównanie efektów działania AI (zwłaszcza gdy mówimy o systemach uczących się) oraz efektów działania zwierząt poprzez odwołanie się do relacji obydwu do człowieka wydaje się uprawnione (o czym w dalszej części monografii).

wych na te, które mogą, i na te, które nie mogą być *de lege lata* uznane za przedmiot prawa autorskiego. Takie dzieła komputerowe, które *de lege lata* mogą być uznane za przejaw działalności twórczej o indywidualnym charakterze, określam w monografii mianem dzieł wspomaganych komputerowo. Te z kolei, które nie mogą spełniać przesłanek ochrony prawnoautorskiej, nazywam dziełami generowanymi komputerowo. Co istotne, przyjętego *criterium divisionis* nie stanowi to, czy w danym obiekcie *in concreto* zaznaczają się cechy twórczości i indywidualności, ale to, czy sposób działania programu komputerowego, który został wykorzystany do jego stworzenia (czyli funkcja programu w procesie kreacyjnym), stwarza możliwość zaznaczenia się tych cech w dziele. Zgodnie z przyjętym w tej części pracy założeniem sposób działania programu implikuje bowiem to, czy w procesie powstawania określonego rezultatu będzie się dawał wyodrębnić taki wkład człowieka, którego zakres pozwoliłby na przypisanie owego rezultatu jego twórczej aktywności. Chodzi tu zatem o potencjalną możliwość spełnienia przesłanek ochrony prawnoautorskiej, a nie o to, czy oceniane dzieło komputerowe rzeczywiście spełnia kryteria ochrony. Zgodnie z tym dychotomicznym podziałem uporządkowane zostały rozważania dalszej części rozdziału trzeciego.

Celem rozdziału czwartego było ustalenie, czy wizualne i dwuwymiarowe dzieła generowane komputerowo, a więc takie, w których nie zaznacza się twórczy wkład człowieka, których obiektywnie obserwowalne cechy uzasadniałyby stwierdzenie, że mamy do czynienia z podlegającym ochronie prawnoautorskiej utworem, mogą podlegać *de lege lata* ochronie na gruncie przepisów polskiego prawa prywatnego w sposób, który uzasadniałby przyjęcie, że stanowią one niematerialne dobra prawne, a także sformułowanie wniosków *de lege ferenda*.

Monografię zamyka zakończenie, w którym przedstawiono wnioski końcowe.

Podstawową metodą badawczą wykorzystaną w niniejszej pracy była metoda dogmatycznoprawna. W opracowaniu wykorzystano pomocniczo elementy badania funkcjonalnego, dla których analiza dogmatyczna stanowiła warunek *sine qua non*. Metoda funkcjonalna wykorzystana została zarówno w wariancie teoretycznym, jak i empirycznym. W tym pierwszym polegała ona na badaniu hipotetycznych scenariuszy wydarzeń, które zostały poddane analizie przypadku. W drugim zaś polegała na wykorzystaniu metod prawnoempirycznych w celu oceny prawnoautorskiego statusu rezultatów stworzonych przeze mnie na potrzeby badania z wykorzystaniem AI oraz w celu analizy wykorzystywanych w obrocie wzorców umów dotyczących korzystania z generatywnych programów AI. Wykorzystanie badań funkcjonalnych miało na celu poszerzenie perspektywy poznawczej, przez wykroczenie poza analizę logiczno-językową i wykładnię tekstów prawnych. W ostatniej części pracy, a więc w jej rozdziale czwartym – w celu sformułowania wniosków *de lege ferenda* – sięgnięto do narzędzi właściwych dla ekonomicznej analizy prawa.

Niniejsze opracowanie uwzględnia stan prawny na dzień 1.01.2025 r.

Rozdział 1

SZTUCZNA INTELIGENCJA

1.1. Wstęp

W niniejszym rozdziale dokonano syntetycznego przeglądu fundamentalnych zagadnień związanych z rozwojem i funkcjonowaniem technologii sztucznej inteligencji. Nie jest bynajmniej moją aspiracją dogłębne omówienie całości problematyki AI ani przedstawienie wszechstronnego przeglądu dostępnej w tym zakresie literatury. Nie wydaje się to ani możliwe, ani konieczne. Zarysowywana w tym rozdziale dziedzina jest bowiem niezmiernie obszerna i obejmuje wiedzę z zakresu różnych dyscyplin nauki. Stąd jej kompleksowe omówienie dalece wykraczałoby poza ramy opracowania.

1.2. Czym jest sztuczna inteligencja?

Odpowiedź na pytanie, czym jest sztuczna inteligencja, nie jest prosta przynajmniej z dwóch powodów. Po pierwsze, jak zauważa J. Kaplan, nie ma powszechnej zgody co do tego, czym w ogóle jest inteligencja. Po drugie, kontynuuje, iż „mamy niewiele powodów, żeby być przekonanymi, że inteligencja maszynowa ma duży związek z inteligencją ludzką, przynajmniej na razie”[1]. Zanim rozwiniemy te dwa argumenty, należy przypomnieć, jak sztuczną inteligencję definiował twórca tego terminu – J. McCarthy. Otóż we wniosku o sfinansowanie letniej konferencji w Dartmouth, przedstawionym w 1955 r. Fundacji Rockefellera, zaprezentował on problematykę sztucznej inteligencji jako ideę stworzenia maszyny zachowującej się w sposób, który nazwalibyśmy inteligentnym, gdyby tak zachowywał się człowiek[2].

---

[1] J. Kaplan, *Sztuczna inteligencja. Co każdy powinien wiedzieć*, tłum. S. Szymański, Warszawa 2019, s. 15.

[2] „For the present purpose the artificial intelligence problem is taken to be that of making a machine behave in ways that would be called intelligent if a human were so behaving” za: J. McCarthy [w:] J. McCarthy, M.L. Minsky, N. Rochester, C.E. Shannon, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, 1955, s. 11, http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf (dostęp: 1.12.2024 r.).

W nauce funkcjonuje wiele konkurencyjnych definicji inteligencji. Jedne odwołują się do kategorii takich jak umiejętność logicznego myślenia, inne kładą nacisk na zdolność przyswajania wiedzy, rozwiązywanie problemów, planowanie, jeszcze inne akcentują takie atrybuty jak samoświadomość czy kreatywność[3]. Wszystkie te definicje łączy jednak jeden wspólny mianownik. Każda z nich operuje swoistą kategorią celu. Inteligencja definiowana jest bowiem każdorazowo przez pryzmat zadania, do realizacji którego zdolny jest oceniany podmiot. Różne cele to w konsekwencji różne rodzaje inteligencji. Z takim wnioskiem koresponduje kontrowersyjna teoria inteligencji wielorakiej H. Gardnera, który wyróżnia osiem jej rodzajów: od „logiczno-matematycznej" przez „społeczną" po „cielesno-kinestetyczną"[4]. Akceptując pogląd o istnieniu różnych rodzajów inteligencji, M. Tegmark proponuje na przykład, by inteligencję definiować możliwie szeroko. Sam ujmuje ją jako „zdolność do osiągania złożonych celów"[5].

Nie mniejsze trudności niż ujęcie inteligencji w jednoznaczne ramy definicyjne przysparza jej pomiar. Przypuszczalnie najpopularniejszym sposobem pomiaru ludzkiej inteligencji jest test IQ. Czy jednak próba kwantyfikacji czegoś z natury rzeczy tak subiektywnego jak inteligencja ma jakikolwiek sens? Nie ulega wątpliwości, że „[n]asza kulturowa predylekcja do redukowania rzeczy do pomiarów numerycznych, które ułatwiają bezpośrednie porównywanie, często tworzy fałszywy pozór obiektywności i precyzji"[6]. Testy IQ ignorują bowiem niektóre istotne sprawności umysłowe, takie jak zdolności artystyczne czy kreatywność[7]. Wspomniana teoria inteligencji wielorakiej H. Gardnera, choć z założenia miała stanowić panaceum na przesadnie upraszczające problematykę inteligencji standardowe metody pomiaru IQ, od ponad 30 lat pozostaje w ogniu krytyki środowiska naukowego[8]. Czy zatem możemy dokonać pomiaru inteligencji maszyn przy zastosowaniu jakiegokolwiek testu wykorzystywanego dziś do mierzenia ludzkiej inteligencji? Według koncepcji M. Tegmarka inteligencję, rozumianą jako zdolność do osiągania złożonych celów, można mierzyć wyłącznie za pomocą spektrum zdolności w odniesieniu do celów szczegółowych[9].

Choć fundamentalnym problemem na drodze do osiągnięcia przez badaczy konsensusu co do tego, czym jest sztuczna inteligencja, wydaje się brak jednoznacznej definicji i uniwersalnych narzędzi pomiaru inteligencji ludzkiej, nie należy zapominać, że wśród badaczy nie ma również zgody co do tego, do jakich wzorców należy się odwoływać przy dokonywaniu oceny, czy dane rozwiązanie może być uznane za AI. Nie wszystkie

[3] M. Tegmark, *Życie 3.0. Człowiek w erze sztucznej inteligencji*, tłum. T. Krzysztoń, Warszawa 2019, s. 72.

[4] H. Gardner, *Frames of Mind: The Theory of Multiple Intelligences*, New York 2011, s. 77 i n.

[5] M. Tegmark, *Życie 3.0*..., s. 72.

[6] J. Kaplan, *Sztuczna inteligencja*..., s. 16.

[7] T. Walsh, *To Żyje! Sztuczna inteligencja. Od logicznego fortepianu po zabójcze roboty*, tłum. W. Sikorski, Warszawa 2018, s. 54.

[8] Zob. M. Tyszkowska, *Metaanaliza krytycznego dyskursu nad teorią wielorakich inteligencji Howarda Gardnera*, „Przegląd Badań Edukacyjnych" 2017/1, s. 137–149.

[9] M. Tegmark, *Życie 3.0*..., s. 73.

bowiem ujęcia sztucznej inteligencji odwołują się do kategorii inteligencji ludzkiej, a te, które to czynią, nie zawsze robią to w ten sam sposób[10].

Odżegnując się od udziału w naukowej debacie dotyczącej miary inteligencji, uważam, że warto zwrócić uwagę na to, w jaki sposób inteligencja bywa kwantyfikowana poza dyskursem naukowym. Nie wydaje się budzić większych kontrowersji konstatacja, że intuicyjnie możemy stwierdzić, iż z określonych względów jedna osoba jest inteligentniejsza od drugiej w danej dziedzinie. Skoro powiemy, że uczeń, który szybciej dodaje i odejmuje ciągi liczb, jest inteligentniejszy od tego, który robi to wolniej, to dlaczego bawi nas sugestia, że prosty kalkulator kieszonkowy jest bardziej inteligentny od każdego człowieka na ziemi? Otóż dla oceny inteligencji maszynowej relewantne jest nie tylko to czy, lecz także jak określony problem został rozwiązany.

Kiedy oceniamy zdolności ludzi w dodawaniu i odejmowaniu, nie badamy, jak rozwiązali oni zadane równania. Czy jednak program komputerowy, który wygrywa partię w kółko i krzyżyk, sprawdzając uprzednio zaprogramowane mu możliwe sekwencje gry (poruszając się po tzw. drzewie gry metodą przeszukiwania przestrzeni stanów), potraktujemy tak samo jak program, który poznaje jej reguły oraz strategie zwycięstwa, obserwując grających ludzi, a następnie rozgrywając liczne partie sam ze sobą? Większość z nas odrzuci zapewne możliwość uznania tego pierwszego za sztuczną inteligencję, będąc skłonnym do przyznania takiego statusu drugiemu z programów.

Skoro jednak, jak dostrzega J. Kaplan, istnieje „praktyczna równoważność między wyborem odpowiedzi z niezmiernie dużej liczby możliwości a intuicyjnym udzieleniem odpowiedzi dzięki wnikliwości i kreatywności”[11], to dlaczego oceniając, czy maszyna jest inteligentna, czynimy przedmiotem naszego zainteresowania to, jak osiągnęła ona dany rezultat. To właśnie wykorzystanie umiejętności, wyuczonych lub wrodzonych, które można zastosować w szerokim spektrum problemów, czyli zdolność do dokonywania właściwych uogólnień na podstawie ograniczonych danych, jest zdaniem J. Kaplana istotą inteligencji[12]. Nie zawsze jest więc przejawem zachowania inteligentnego takie działanie maszyny, dzięki któremu uzyskuje ona taki sam lub lepszy od człowieka rezultat w realizacji określonego zadania. Natomiast „[i]m szersza dziedzina zastosowania, im szybsze wnioski wyciąga się z minimalnej informacji, tym bardziej inteligentne jest zachowanie. Jeśli ten sam program, który uczy się gry w kółko i krzyżyk może nauczyć się każdej gry planszowej, tym lepiej”[13].

[10] Zamiast wielu zob. S.J. Russell, P. Norvig, *Artificial Intelligence. A Modern Approach*, Upper Saddle River, New Jersey 2010, s. 2.

[11] J. Kaplan, *Sztuczna inteligencja*..., s. 18.

[12] J. Kaplan, *Sztuczna inteligencja*..., s. 21.

[13] J. Kaplan, *Sztuczna inteligencja*..., s. 21.

1.3. Dzieje sztucznej inteligencji

1.3.1. Pierwsze sukcesy i porażki

Niektóre z problemów, które dziś uznalibyśmy za należące do dziedziny sztucznej inteligencji, nurtowały filozofów już od czasów starożytnych[14]. Marzenie o zbudowaniu myślącej maszyny na dobre zagościło w myśli naukowej w pierwszej połowie XX w., jednak dopiero pojawienie się elektronicznych urządzeń obliczeniowych po II wojnie światowej umożliwiło przekucie założeń teoretycznych w rzeczywistość. Impulsem do tego miał stać się *Dartmouth Summer Research Project on Artificial Intelligence*.

> „Proponujemy przeprowadzenie dwumiesięcznego studium obejmującego dziesięciu badaczy sztucznej inteligencji [...]. Podjęta zostanie próba odkrycia, jak zbudować maszyny posługujące się językiem naturalnym, formułujące abstrakcje i idee, rozwiązujące problemy zarezerwowane dziś dla ludzi i doskonalące same siebie. Sądzimy, że można dokonać znaczących postępów w jednej lub kilku z tych dziedzin, jeśli starannie wyselekcjonowana grupa naukowców podejmie nad nimi wspólną pracę w ciągu lata"[15].

Ta śmiała prognoza pochodzi z przedstawionego Fundacji Rockefellera wniosku o sfinansowanie konferencji w Dartmouth i doskonale oddaje stan ducha jej pomysłodawców i uczestników. Latem 1956 r. w konferencji udział wzięło liczne grono prominentnych naukowców, z których wielu zostało później uznanych za ojców sztucznej inteligencji. Choć nie jest jasne, co (jeśli cokolwiek w ogóle) osiągnięto podczas konferencji w Dartmouth, wydarzenie to przeszło do historii jako impuls do wykształcenia się sztucznej inteligencji jako odrębnej dziedziny badań naukowych[16]. Przypuszczalnie najważniejszymi rezultatami tego wydarzenia było nawiązanie relacji otwierających drzwi do dalszej wymiany wiedzy między najwybitniejszymi badaczami i pionierami w dziedzinie inteligentnych maszyn[17] oraz niewiarygodny sukces, jaki termin „sztuczna inteligencja" odniósł „w przyciąganiu zainteresowania i uwagi sięgających daleko poza jego akademickie korzenie"[18].

Ekscytacja i optymizm inicjatorów konferencji w Dartmouth odnośnie do tego, co i jak szybko zostanie osiągnięte w dziedzinie sztucznej inteligencji, nie po raz ostatni okazały się jednak przesadne. Od tego czasu w badaniach nad sztuczną inteligencją dały się zaobserwować naprzemienne okresy wzmożonego zainteresowania i wielkich nadziei

---

[14] M. Flasiński, *Wstęp do sztucznej inteligencji*, Warszawa 2018, s. 3; I. Goodfellow, Y. Bengio, A. Courville, *Deep learning. Systemy uczące się*, tłum. W. Sikorski, Warszawa 2018, s. 1.

[15] J. McCarthy, M.L. Minsky, N. Rochester, C.E. Shannon, *A Proposal for the Dartmouth Summer Research Project*..., s. 2.

[16] N. Bostrom, *Superinteligencja. Scenariusze, strategie, zagrożenia*, tłum. D. Konowrocka-Sawa, Warszawa 2016, s. 23.

[17] S.J. Russell, P. Norvig, *Artificial Intelligence*..., s. 18.

[18] J. Kaplan, *Sztuczna inteligencja*..., s. 33.

oraz niepowodzeń i pesymizmu, tzw. zim sztucznej inteligencji. Historię badań nad AI można by więc w uproszczeniu przedstawić, posługując się ilustracją sinusoidy, w której wyraźnie zaznaczają się cykle hossy – optymizmu i zwiększonego finansowania, oraz bessy – rozczarowań i ograniczonych funduszy[19].

Chociaż początkowe postępy w dziedzinie sztucznej inteligencji nie były ani tak szybkie, ani tak spektakularne, jak przewidywali uczestnicy konferencji w Dartmouth, wiele projektów zrealizowanych w tym okresie uznaje się dziś za kamienie milowe w rozwoju inteligentnych maszyn. Sceptycyzm przedstawicieli naukowego establishmentu mobilizował badaczy sztucznej inteligencji, a każde twierdzenie typu „maszyna nigdy nie zdoła zrobić X" było niczym rzucona rękawica. Naturalnie pionierzy sztucznej inteligencji odpowiadali, demonstrując programy, którym udawało się owo X zrobić[20]. Systemy powstające w tamtym czasie koncentrowały się więc na prostych problemach (ang. *toy problems*), które mogły zostać rozwiązane w ograniczonej, dobrze zdefiniowanej dziedzinie, umożliwiającej zademonstrowanie zredukowanej wersji określonej umiejętności (w tzw. mikroświecie)[21], takiej jak gra w różnego rodzaju gry[22].

Do wczesnych przełomów w dziedzinie sztucznej inteligencji należy system Logic Theorist, stworzony w 1956 r. przez A. Newella i H.A. Simona, który dowiódł większości twierdzeń zawartych w książce *Principia Mathematica* autorstwa A.N. Whiteheada i B. Russella z 1910 r., demonstrując tym samym, że maszyny są zdolne do przeprowadzania rozumowania dedukcyjnego[23]. General Problem Solver, skonstruowany kilka lat później przez ten sam zespół, wykazywał zdolność rozwiązywania rozmaitych formalnie opisanych problemów[24]. Rezultatem metodologicznym badań A. Newella i H.A. Simona było sformułowanie podejścia do konstrukcji inteligentnych systemów, zwanego symulacją kognitywną, zgodnie z którym system powinien symulować ludzkie procesy umysłowo-poznawcze[25].

W połowie lat 60. XX w. liczne laboratoria prowadzące w Stanach Zjednoczonych badania nad sztuczną inteligencją uzyskały potężne wsparcie finansowe ze strony Agencji Zaawansowanych Projektów Badawczych w Obszarze Obronności (ang. Defence Advanced Research Projects Agency, DARPA)[26]. Jednym z beneficjentów finansowania DARPA było laboratorium badawcze Stanford Research Institute (od 1977 r.: SRI Inter-

[19] N. Bostrom, *Superinteligencja*..., s. 23.

[20] S.J. Russell, P. Norvig, *Artificial Intelligence*..., s. 18.

[21] S.J. Russell, P. Norvig, *Artificial Intelligence*..., s. 19.

[22] J. Kaplan, *Sztuczna inteligencja*..., s. 35; N. Bostrom, *Superinteligencja*..., s. 23–24.

[23] A. Newell, H.A. Simon, *The logic theory machine – A complex information processing system*, 15.06.1956 r., https://exhibits.stanford.edu/feigenbaum/catalog/ct530kb5673 (dostęp: 1.12.2024 r.).

[24] A. Newell, H.A. Simon, *GPS: A Program That Simulates Human Thought* [w:] *Lernende automaten*, red. H. Billing, München 1961, s. 109–110.

[25] M. Flasiński, *Wstęp*..., s. 16 i n.

[26] J. Kaplan, *Sztuczna inteligencja*..., s. 36.

national), gdzie w latach 1966–1972 realizowano projekt, którego celem było skonstruowanie pierwszego „autonomicznego" robota. Shakey, owoc prac zespołu badawczego SRI, zaprezentował, w jaki sposób mogą zostać zintegrowane osiągnięcia w zakresie rozumowania logicznego z widzeniem komputerowym, mapowaniem, planowaniem i wreszcie fizycznym działaniem[27].

Głośnym echem w świecie nauki odbiło się stworzenie pierwszego programu rzekomo „rozumiejącego" język naturalny. SHRDLU, system stworzony w 1970 r. przez T. Winograda, zademonstrował, że zasymulowane ramię robota w wirtualnym świecie geometrycznych bloków może wykonywać komendy i odpowiadać na pytania wprowadzone na klawiaturze w języku naturalnym. SHRDLU połączył „postępy w analizie składni, semantyce, odpowiadaniu na pytania, dialogu, logice, reprezentacji wiedzy i grafice komputerowej", niosąc za sobą milczącą obietnicę, że „niebawem możemy być w stanie wchodzić w dialog z komputerami jako równymi nam pod względem intelektualnym"[28]. System budził wiele emocji i ogromne nadzieje, lecz nie należy zapominać, że działanie SHRDLU ograniczało się do zamkniętego mikroświata bloków[29]. Badacze sztucznej inteligencji zakładali wówczas, że rozumienie języka naturalnego wymagałoby w istocie rzeczy, by maszyna posiadała ogólną wiedzę o świecie[30].

W połowie lat 60. ubiegłego wieku pojawił się nowy kierunek badań, podobnie jak symulacja kognitywna należący do nurtu symbolicznej sztucznej inteligencji, który zapoczątkował nowy trend w dziedzinie inteligentnych systemów – tzw. podejście oparte na wiedzy[31]. Na podstawie tego paradygmatu konstruowano systemy zwane eksperckimi. Narodziny podejścia opartego na wiedzy są związane z projektem Dendal, realizowanym przez E. Feigenbauma i J. Lederberga od 1965 r., który był śmiałą próbą zakodowania wiedzy eksperckiej (w tym przypadku wiedzy z zakresu chemii molekularnej) w programie komputerowym[32]. Systemy te, zazwyczaj bardzo wąsko wyspecjalizowane, wyciągały proste wnioski z bazy wiedzy tworzonej przez programistów, nazywanych „inżynierami wiedzy", którą ci pozyskali od ekspertów w danej dziedzinie, a następnie przekazywali ją do systemu w postaci sformalizowanej reprezentacji[33].

Nieadekwatność metod, które zapewniły powodzenie wczesnym wersjom programów i wyzwoliły pierwszą serię sukcesów, do szerszego zakresu problemów lub problemów o większym stopniu trudności okazała się bolączką badaczy sztucznej inteligencji

[27] SRI International, *Shakey the robot*, https://www.sri.com/hoi/shakey-the-robot/ (dostęp: 1.12.2024 r.).

[28] J. Kaplan, *Sztuczna inteligencja*..., s. 37.

[29] S.J. Russell, P. Norvig, *Artificial Intelligence*..., s. 23.

[30] S.J. Russell, P. Norvig, *Artificial Intelligence*..., s. 23.

[31] M. Flasiński, *Wstęp*..., s. 5–6.

[32] T. Walsh, *To Żyje!*..., s. 36.

[33] M. Flasiński, *Wstęp*..., s. 6.

w połowie lat 70. XX w.[34] Popularne w owym czasie metody przeszukiwania przestrzeni stanów[35] oparte na symulacji kognitywnej dobrze sprawdzały się w odniesieniu do relatywnie prostych problemów, wymagających sformułowania serii kroków w celu osiągnięcia określonego rezultatu. Metody te nie pozwalały jednak na przezwyciężenie tzw. eksplozji kombinatorycznej, a więc gwałtownego wzrostu możliwych sekwencji kroków. Stawienie czoła temu wyzwaniu wymagało zastosowania bardziej zaawansowanych metod przeszukiwania, choćby wzmocnionych rozumowaniem heurystycznym (wykorzystujących funkcję heurystyczną, która pozwala szacować odległość każdego z wierzchołków grafu od rozwiązania, tym samym redukując przestrzeń przeszukiwania[36]), które jednak w tamtym czasie nie były jeszcze wystarczająco rozwinięte. Dziedzina sztucznej inteligencji napotkała w tym okresie także na dwa inne problemy. Barierą dla dalszego rozwoju okazały się bowiem niedostateczna ilość danych oraz istotne ograniczenia sprzętowe, gdy chodzi o pamięć i moc obliczeniową ówczesnych komputerów[37].

Rosnący sceptycyzm zaowocował ograniczeniem funduszy przeznaczanych na badania nad AI i doprowadził do tzw. pierwszej zimy sztucznej inteligencji. Odwilż nastąpiła dopiero na początku lat 80., kiedy do wyścigu technologicznego, w którym stawką była pozycja lidera komputeryzacji, dołączyła Japonia. W 1981 r. w kraju tym uruchomiono hojnie finansowany projekt systemów komputerowych piątej generacji (ang. *Fifth Generation Computer Systems*). Ambicje Japonii sprowokowały zachodnich rywali, w szczególności Stany Zjednoczone oraz Wielką Brytanię, do pójścia w jej ślady. Jednak ani projektowi piątej generacji, ani siostrzanym projektom realizowanym na Zachodzie nie udało się osiągnąć zamierzonych celów[38]. W latach 80. doszło również do upowszechnienia się wspomnianych już systemów eksperckich. Choć, jak już była o tym mowa, podejście to narodziło się wcześniej, bo już w połowie lat 60., to w początkowym okresie nie cieszyło się popularnością z uwagi na niewielką ilość wiedzy dostępnej w postaci cyfrowej. Na początku lat 80. produkty i usługi o charakterze systemów eksperckich przyciągnęły jednak uwagę kapitału inwestycyjnego, doprowadzając do gwałtownego wzrostu koniunktury w dziedzinie sztucznej inteligencji[39]. Systemy te okazały się jednak bardzo kłopotliwe w użyciu, a ich rozwój, aktualizacja i weryfikacja – niezwykle kosztowne[40]. Wkrótce nastała „druga zima sztucznej inteligencji".

[34] S.J. Russell, P. Norvig, *Artificial Intelligence*..., s. 20 i n.

[35] Przestrzeń stanów to graf, którego wierzchołki reprezentują możliwe do wykonania kroki w procesie rozwiązywania problemu. Rozwiązanie problemu sprowadza się w tym przypadku do znalezienia na grafie ścieżki od wierzchołka początkowego do wierzchołka końcowego. Zob. M. Flasiński, *Wstęp*..., s. 35–39.

[36] M. Flasiński, *Wstęp*..., s. 43–46.

[37] S.J. Russell, P. Norvig, *Artificial Intelligence*..., s. 21 i n.

[38] S.J. Russell, P. Norvig, *Artificial Intelligence*..., s. 24.

[39] J. Kaplan, *Sztuczna inteligencja*..., s. 43.

[40] N. Bostrom, *Superinteligencja*..., s. 26.

1.3.2. Droga ku głębokiemu uczeniu

W połowie lat 80. XX w. uwaga badaczy sztucznej inteligencji zwróciła się w kierunku nowej technologii. Tak zwane sztuczne sieci neuronowe (ang. *Artificial Neural Network*, ANN) miały stanowić alternatywę względem tradycyjnych podejść opartych na systemach symbolicznych, m.in. symulacji kognitywnej i podejścia opartego na wiedzy (warto wskazać, że do kategorii podejść opartych na systemach symbolicznych należą też m.in. podejście oparte na logice, podejście statystyczne[41]). Nowe techniki, uznawane dziś za należące do nurtu subsymbolicznej AI, chlubiły się działaniem bardziej organicznym i dawały nadzieję na przezwyciężenie problemów, z jakimi zmagały się systemy należące do trwającego w zastoju nurtu symbolicznej AI, znanego dziś jako „stara dobra sztuczna inteligencja" (ang. *Good Old Fashioned Artificial Intelligence*, GOFAI)[42].

1.3.2.1. Sieci neuronowe

Neuronaukowa perspektywa w badaniach nad sztuczną inteligencją opiera się na założeniu, że ludzki mózg stanowi niepodważalny dowód na to, iż inteligentne zachowanie jest możliwe, a najprostszą drogą do jego odtworzenia w postaci komputerowej jest algorytmiczne zasymulowanie zasad działania mózgu biologicznego. Stąd pierwsze algorytmy uczenia się, motywowane perspektywą neuronaukową, miały na celu komputerowe modelowanie biologicznych procesów uczenia się[43]. Pierwszym matematycznym modelem funkcji mózgu był model neuronu biologicznego opracowany przez W.S. McCullocha oraz W. Pittsa w 1943 r.

Proste sieci neuronowe, budowane na podstawie tzw. neuronu McCullocha-Pittsa, znane były już w latach 50. minionego stulecia[44]. ANN inspirowane są strukturą i funkcjonowaniem ich biologicznych pierwowzorów – sieci neuronalnych w mózgu. Nie są one jednak projektowane jako realne modele sieci biologicznych. Niektórzy badacze podejmują próby ścisłego odwzorowania sieci połączeń neuronalnych w mózgu poprzez stworzenie kompletnej mapy tych połączeń, zwanej konektomem, i jej odtworzenie w postaci programu komputerowego. Tak rozumiana emulacja nie jest jednak przedmiotem zainteresowania głównego nurtu badań[45]. Choć przy okazji relacjonowania przełomowych osiągnięć ANN, zarówno tych pionierskich, jak i współczesnych, często akcentuje się związek pomiędzy tymi sieciami a ich biologicznymi pierwowzorami, należy podkreślić, że sztucznych sieci neuronowych nie powinno się postrzegać jako

[41] Por. A. Przegalińska, P. Oksanowicz, *Sztuczna inteligencja. Nieludzka, arcyludzka*, Kraków 2020, s. 46 i n.

[42] N. Bostrom, *Superinteligencja*..., s. 26.

[43] I. Goodfellow, Y. Bengio, A. Courville, *Deep learning*..., s. 13–14.

[44] M. Flasiński, *Wstęp*..., s. 161.

[45] J. Kaplan, *Sztuczna inteligencja*..., s. 49.

próby symulacji mózgu[46]. Nowoczesne modele ANN czerpią inspirację z wielu dziedzin, wśród których I. Goodfellow, Y. Bengio, A. Courville wymieniają m.in. algebrę liniową, prawdopodobieństwo czy optymalizację numeryczną[47]. Analizując przyczyny ograniczania roli neuronauki w obecnych badaniach nad sztuczną inteligencją, wskazani autorzy zwracają uwagę na to, że nie osiągnięto jeszcze wystarczającego stopnia zrozumienia algorytmów biologicznych, na których opiera się działanie ludzkiego mózgu[48].

Jednym z prekursorów podejścia opartego na ANN był F. Rosenblatt, który w drugiej połowie lat 50. stworzył perceptron: prostą sieć neuronową i protoplastę sieci głębokich, implementującą algorytm uczący, przeznaczony do klasyfikowania obrazów[49]. Do rychłego porzucenia tego obszaru badań niewątpliwie przyczynił się traktat matematyczny Perceptrons, napisany w 1969 r. przez M. Minsky'ego i S. Paperta, w którym naukowcy wykazali, że możliwości perceptronów jednowarstwowych są bardzo ograniczone[50]. Choć rozważali oni możliwość zastąpienia modeli jednowarstwowych sieciami wielowarstwowymi, wyrazili wątpliwości co do tego, czy sieci składające się z większej ilości warstw kiedykolwiek powstaną. Sceptycyzm ten udzielił się licznym przedstawicielom środowiska naukowego, w konsekwencji czego podejście oparte na ANN straciło na popularności.

ANN przeżyły swój renesans pod koniec lat 80., przerywając tym samym „drugą zimę AI". Druga fala badań nad ANN powszechnie kojarzona jest z koneksjonistycznym podejściem do sztucznej inteligencji, które opiera się na założeniu, że uczenie jest wynikiem asocjacji powstającej między bodźcem a odpowiedzią na ten bodziec[51]. Należy jednak wyjaśnić, że w tzw. modelach koneksjonistycznych asocjacje mogą być reprezentowane za pomocą dwóch odmiennych typów sieci[52]. Pierwszy z nich stanowią sieci koneksjonistyczne z reprezentacją lokalną, w których każdy składnik wiedzy zapamiętywany jest przez pojedynczy element sieci. Do tej kategorii należą np. sieci semantyczne, zaliczane do nurtu symbolicznej AI[53]. Drugim typem sieci koneksjonistycznych są tzw. sieci z reprezentacją rozproszoną, w których wiedza dystrybuowana jest pomiędzy wiele elementów sieci[54]. Zgodnie z fundamentalną ideą tak pojmowanego koneksjonizmu „duża liczba prostych elementów obliczeniowych może razem w sieci osiągnąć inteligentne zachowanie"[55]. ANN powszechnie kojarzone z koneksjonistycznym podejściem do sztucznej inteligencji należą do drugiego ze wskazanych typów.

[46] I. Goodfellow, Y. Bengio, A. Courville, *Deep learning*..., s. 16.

[47] I. Goodfellow, Y. Bengio, A. Courville, *Deep learning*..., s. 16.

[48] I. Goodfellow, Y. Bengio, A. Courville, *Deep learning*..., s. 15.

[49] T.J. Sejnowski, *Deep Learning. Głęboka rewolucja. Kiedy sztuczna inteligencja spotyka się z ludzką*, tłum. P. Cypryański, Warszawa 2019, s. 57 i n.

[50] T.J. Sejnowski, *Deep Learning*..., s. 65 i n.

[51] M. Flasiński, *Wstęp*..., s. 26–27.

[52] M. Flasiński, *Wstęp*..., s. 26–27.

[53] M. Flasiński, *Wstęp*..., s. 96 i n.

[54] M. Flasiński, *Wstęp*..., s. 96 i n.

[55] I. Goodfellow, Y. Bengio, A. Courville, *Deep learning*..., s. 17.

Rozproszona reprezentacja wspomagana przez inne istotne osiągnięcie lat 80. – algorytm wstecznej propagacji błędu D. Rumelharta[56] – stała się jednym z motorów nowej „wiosny sztucznej inteligencji". Pod koniec XX w. dowiedziono tym samym, że – wbrew dominującej dotychczas opinii – stworzenie sieci wielowarstwowej i przezwyciężenie ograniczeń perceptronu było możliwe. Druga fala badań nad ANN trwała do połowy lat 90., kiedy to w poszukiwaniu funduszy po raz kolejny zaczęto brnąć w nierealistyczne obietnice. Pierwsze wielowarstwowe sieci neuronowe uznawano za trudne do nauczenia, a zawiedzeni brakiem spektakularnych sukcesów inwestorzy po niedługim czasie wycofali się z finansowania projektów związanych z rozwojem ANN[57]. U progu drugiego tysiąclecia gorączka sieci neuronowych wydawała się opaść.

Trzecia fala badań rozpoczęła się wraz z nadejściem drugiej dekady XXI w. i spopularyzowała tzw. głębokie sieci neuronowe (ang. *Deep Neural Network*, DNN), określane również jako głębokie sieci jednokierunkowe lub wielowarstwowe perceptrony[58]. Stały się one siłą napędową wielu przełomowych osiągnięć w zakresie sztucznej inteligencji i tworzą dziś podstawę wielu komercyjnych zastosowań.

1.3.2.2. Świt głębokiego uczenia

U źródła popularności ANN legło przekonanie, że czerpanie z bogatego dorobku natury jest dobrą zasadą metodologiczną dla konstrukcji systemów sztucznie inteligentnych[59]. Czynności, takie jak: widzenie, słyszenie, mówienie czy poruszanie się przychodzą nam wszak z łatwością. Wydawać by się więc mogło, że skonstruowanie programu posiadającego umiejętności właściwe inteligentnym organizmom biologicznym będzie względnie proste. Pionierzy sztucznej inteligencji przekonali się jednak, że ocena zadań ludzką miarą może dawać mylący obraz w odniesieniu do stopnia ich trudności dla możliwości komputerów.

Jak pisał na początku XXI w. psycholog S. Pinker, główną lekcją wyniesioną z badań nad AI jest to, że „trudne problemy są łatwe, a łatwe problemy są trudne"[60]. Należy bowiem zauważyć, że moc obliczeniowa niezbędna do przeprowadzenia wysokopoziomowego rozumowania jest niewielka, podczas gdy naturalne atrybuty ludzkiego mózgu, takie jak zdolności sensomotoryczne i percepcja wymagają olbrzymich zasobów obliczeniowych[61]. W konsekwencji stosunkowo łatwo zaprogramować komputery tak, by wykonywały skomplikowane działania matematyczne czy nawet posiadły umiejętność gry w szachy na poziomie przewyższającym zdolności człowieka, a niezwykle trudno

[56] T.J. Sejnowski, *Deep Learning*..., s. 137 i n.
[57] I. Goodfellow, Y. Bengio, A. Courville, *Deep learning*..., s. 18.
[58] I. Goodfellow, Y. Bengio, A. Courville, *Deep learning*..., s. 165.
[59] M. Flasiński, *Wstęp*..., s. 13.
[60] S. Pinker, *The Language Instinct. How the Mind Creates Language*, New York 2007, s. 190–191.
[61] M. Tegmark, *Życie 3.0*..., s. 75–76.

sprawić, aby przejawiały zdolność rozpoznawania obrazów odpowiadającą możliwościom kilkuletniego dziecka. Zależność ta, znana jako paradoks Moraveca, sformułowana została w latach 80. XX w. przez H. Moraveca, R. Brooksa i M. Minsky'ego. M. Minsky podkreślał przy tym, że najtrudniejsze do zaprogramowania są te umiejętności, których sobie nie uświadamiamy. „Generalnie, jesteśmy najmniej świadomi tego, co nasze umysły robią najlepiej" – pisał. „Jesteśmy bardziej świadomi prostych procesów, które nie działają dobrze, niż tych złożonych, które działają bezbłędnie"[62]. Do tych ostatnich należą umiejętności niezbędne dla przetrwania gatunku, takie jak widzenie, rozpoznawanie języka czy sensomotoryka. Ich opanowanie wymagało jednak milionów lat ewolucji. „W wysoko rozwiniętych ośrodkach zmysłowych i motorycznych ludzkiego mózgu zakodowane jest miliard lat doświadczenia o naturze świata i o tym, jak w nim przetrwać. [...] Myśl abstrakcyjna jednak jest nową sztuczką, być może młodszą niż sto tysięcy lat. Nie opanowaliśmy jej jeszcze dobrze. Nie jest ona w istocie trudna – taka się tylko wydaje, gdy my ją wykonujemy"[63].

Zdolność rozumowania wymaga wiedzy o tym, w jaki sposób drogą logicznych sądów dojść do ścisłej konkluzji[64]. Wiedza ta w stosunkowo łatwy sposób daje się zaprogramować. Zdobycie umiejętności niezbędnych dla przetrwania gatunku wymaga natomiast umiejętności przeprowadzania generalizacji na podstawie wcześniejszych doświadczeń[65], a więc wymaga zdolności uczenia się. Umiejętność uczenia się uchodzi za jeden z najbardziej fascynujących aspektów inteligencji ogólnej. M. Tegmark podkreśla w tym kontekście, że „historia nauki jest co najmniej tak długa jak historia samego życia, ponieważ każdy samoreprodukujący się organizm kopiuje i przetwarza informacje – zachowania, które zostały w jakiś sposób nauczone"[66]. Nadanie komputerom zdolności uczenia się, jak rozwiązać określony problem na podstawie danych bez wcześniejszego ich zaprogramowania w tym celu, okazało się kluczem do sukcesu systemów „inteligentnych" w dziedzinach dotychczas wymykających się badaczom AI.

Należy przypomnieć, że historycznie pierwszym kierunkiem badań w dziedzinie AI były podejścia symboliczne, m.in. wspomniane już podejście symulacji kognitywnej czy podejście oparte na wiedzy. Do wspólnych założeń metodologicznych podejść należących do nurtu symbolicznej sztucznej inteligencji należy teza, że wiedza powinna być reprezentowana w sposób symboliczny, a wszelkie inteligentne działania można opisać przy użyciu formalnych operacji[67]. Rzeczywiście we wczesnym okresie badań nad AI z powodzeniem rozwiązywano problemy abstrakcyjne, które nadawały się do opisania w sposób sformalizowany. Formalne opisywanie przez programistów wiedzy potrzebnej komputerowi trudno jednak uznać za rozwiązanie optymalne przynajmniej

[62] M. Minsky, *The Society of Mind*, New York 1988, s. 29.
[63] M. Minsky, *The Society*..., s. 15–16.
[64] T.J. Sejnowski, *Deep Learning*..., s. 47.
[65] T.J. Sejnowski, *Deep Learning*..., s. 47.
[66] M. Tegmark, *Życie 3.0*..., s. 105.
[67] M. Flasiński, *Wstęp*..., s. 16.

z kilku powodów. Po pierwsze, twórcy programów komputerowych nie są w stanie dostarczyć systemowi pełni wiedzy o świecie rzeczywistym. Nawet gdyby zaangażować odpowiednio dużą liczbę specjalistów, którzy poświęciliby dekady formalizowaniu i wprowadzaniu do systemu dostępnej wiedzy o świecie, ich praca okazałaby się niewiele warta. Stworzona w ten sposób baza wiedzy nie tylko byłaby w sposób oczywisty niepełna, lecz wymagałaby także nieustannej aktualizacji. Dlatego też, jak już wspomniano, programy uznawane za pierwsze sukcesy w dziedzinie AI ograniczały swoje zastosowanie do względnie sformalizowanych mikroświatów. Trudności te dość szybko stały się dla badaczy AI ewidentne. Stąd już pionierzy tej dziedziny dostrzegli, że konieczne jest wyposażenie systemów w zdolność do przyswajania wiedzy, wyodrębnionej jako pewne wzorce ze zbioru „surowych" danych (ang. *raw data*).

Systemy uczące się doskonalą się dzięki własnym doświadczeniom i zdobytej dzięki nim wiedzy. Jednym z pierwszych systemów uczących się był program do gry w warcaby, stworzony w 1959 r. przez A. Samuela. W tym samym roku A. Samuel ukuł termin „uczenie maszynowe" (ang. *machine learning*). Zarówno pojęcie uczenia maszynowego, jak i stojąca za nim idea zostały spopularyzowane w latach 60. Klasyczne systemy uczące się uznaje się za wpisujące w dominujące we wczesnym okresie podejście symboliczne[68]. Rozwijany od lat 60. prosty algorytm samouczenia zwany naiwnym klasyfikatorem bayesowskim (ang. *naive Bayes classifier*) potrafi np. ocenić, czy otrzymane wiadomości e-mail są chcianą pocztą czy spamem. Działanie klasycznych algorytmów uczących się jest ściśle związane z reprezentatywnością otrzymanych danych. Odpowiednie fragmenty tekstu ocenianej wiadomości e-mail określane były jako cechy. Program uczył się określać, do jakiej kategorii (spam czy niespam) one należą, ale nie mógł wpływać na sposób ich definiowania.

Skuteczność algorytmów uczenia maszynowego silnie związana jest ze sposobem reprezentacji danych, które są im dostarczane. Choć wiele problemów można rozwiązać, projektując odpowiedni dla określonego problemu zestaw cech, a następnie dostarczając je prostemu algorytmowi uczącemu się, dla szerokiego wachlarza zadań trudno określić relewantne cechy. Załóżmy, że chcemy stworzyć program do wykrywania ludzkich twarzy na obrazach. Ludzie rzadko będą mieć trudności z rozpoznaniem danego obiektu na ilustracji jako ludzkiej twarzy, mimo że może ona się różnić pozycją czy położeniem (przeciwnie, obserwowane w nauce zjawisko neurologiczne zwane pareidolią sprawia, że często upatrujemy twarzy w zupełnie przypadkowych elementach świata zewnętrznego).

Zupełnie inaczej jest w przypadku systemów komputerowych. Elementy twarzy, takie jak: oczy, nos, usta czy uszy mogą być co prawda opisane jako cechy za pomocą kształtu i wzajemnego położenia, jednak wszelkie zakłócenia w obrazie twarzy, np. częściowe przesłonięcie czy nietypowy kąt ujęcia, stanowią niejednoznaczności uniemożliwiające programowi identyfikację relewantnych cech. Jednym z rozwiązań tego problemu jest

[68] J. Kaplan, *Sztuczna inteligencja*..., s. 47.

tzw. uczenie się przez reprezentację, tzn. konstrukcja systemów uczących się nie tylko odwzorowywania reprezentacji na wyniki, lecz także samodzielnego formułowania reprezentacji[69]. Trudno zaprzeczyć, że opanowanie przez program reprezentacji będzie rozwiązaniem bardziej wydajnym niż jej ręczne wprowadzanie przez człowieka. Uczenie się przez reprezentację umożliwia też systemom AI szybką adaptację do nowych zadań[70].

Źródłem trudności jest jednak fakt, że na każdy element danych, które możemy obserwować, wpływa wiele czynników zmian. Warunkuje to konieczność rozpoznania czynników zmian i odrzucenia tych, na których nam nie zależy[71]. Wybranie takich cech z nieprzetworzonych danych może być jednak bardzo trudne. Tę fundamentalną trudność uczenia reprezentatywnego rozwiązuje tzw. uczenie głębokie (ang. *deep learning*), które umożliwia programowi zbudowanie złożonych pojęć na podstawie prostszej reprezentacji[72]. Wzorcowym przykładem architektury głębokiego uczenia są wspomniane DNN. W modelu uczenia głębokiego skomplikowane odwzorowanie dzielone jest na serię prostszych, z których każde opisane jest w innej warstwie modelu. „Dane wejściowe są prezentowane w widocznej warstwie, którą nazywamy tak, gdyż zawiera zmienne, które można obserwować. Potem następuje ciąg warstw ukrytych, wyciągających z obrazu coraz bardziej abstrakcyjne cechy. Nazywa się je ukrytymi, gdyż ich wartości nie są podane w danych; model musi sam określić, które pojęcia są użyteczne do wytłumaczenia związków w obserwowanych danych"[73]. System uczenia głębokiego może budować złożone pojęcia opisujące obraz przez połączenie pojęć prostszych.

Rozumowanie symboliczne nie sprawdza się ponadto w przypadku problemów, których reprezentacja w kategoriach symbolicznych jest niemożliwa lub szczególnie utrudniona. Wiedza może bowiem przybierać formy, które nie dają się skodyfikować w symbolicznej formie[74]. Załóżmy, że chcemy skonstruować program komputerowy grający w szachy na poziomie ludzkim. Szachy to stosunkowo prosty mikroświat, obejmujący 64 pola i 32 elementy, które mogą poruszać się według ustalonych zasad. Grę można zatem opisać za pomocą krótkiej listy sformalizowanych reguł[75]. Przyjmijmy jednak, że chcemy skonstruować robota, który potrafi grać w Jengę[76] – grę wymagającą nie tylko znajomości zasad, zręczności, lecz także zrozumienia podstawowych praw fizyki. Z pewnością możemy znaleźć osoby grające w Jengę na poziomie mistrzowskim, ale charakter ich wiedzy eksperckiej, w dużym zakresie intuicyjnej i opartej na typowo ludzkim wyczuciu,

[69] I. Goodfellow, Y. Bengio, A. Courville, *Deep learning*..., s. 4.
[70] I. Goodfellow, Y. Bengio, A. Courville, *Deep learning*..., s. 4.
[71] I. Goodfellow, Y. Bengio, A. Courville, *Deep learning*..., s. 5.
[72] I. Goodfellow, Y. Bengio, A. Courville, *Deep learning*..., s. 5.
[73] I. Goodfellow, Y. Bengio, A. Courville, *Deep learning*..., s. 6.
[74] J. Kaplan, *Sztuczna inteligencja*..., s. 59.
[75] I. Goodfellow, Y. Bengio, A. Courville, *Deep learning*..., s. 2.
[76] Gra zręcznościowa, która polega na tym, że gracze kolejno usuwają ze zbudowanej z drewnianych klocków wieży dowolny klocek i układają go na szczycie budowli. Przegrywa osoba, która wyciągając klocek, przewróci wieżę.

nie daje się łatwo opisać w formie symbolicznej. Dzięki uczeniu głębokiemu systemy komputerowe mogą posiadać wiedzę, którą nie zawsze da się im przekazać językiem sformalizowanym[77].

1.3.2.3. Uczenie nadzorowane i nienadzorowane

Algorytmy uczenia maszynowego mogą być szkolone na kilka sposobów. Wśród podstawowych metod trenowania algorytmów wymienia się uczenie nadzorowane oraz uczenie nienadzorowane. Uczenie nadzorowane (ang. *supervised learning*), czyli tzw. uczenie z nauczycielem, przebiega w sposób, który można porównać do tego, w jaki sposób przebiega nauka dziecka w szkole. W uczeniu nadzorowanym przekazywane algorytmowi dane uczące zawierają dołączone rozwiązania problemu – tzw. etykiety (ang. *labels*)[78]. Algorytmy nadzorowanego uczenia poznając szkoleniowy zbiór danych i powiązane z danymi wyniki, uczą się udzielania prawidłowych odpowiedzi[79]. Oznacza to, że algorytmowi prezentuje się np. zdjęcia przedstawiające ludzkie twarze i takie, na których twarzy nie ma, określając przy tym, które z fotografii zawierają podobiznę człowieka.

Można jednak ograniczyć się do dostarczenia algorytmowi zbioru danych bez wskazywania prawidłowych rozwiązań. Taki sposób uczenia, w którym algorytm samodzielnie uczy się rozpoznawać użyteczne właściwości struktury zbioru danych szkoleniowych i opracowywać funkcję ich przetwarzania (uporządkowywania, wykrywania regularności, klasyfikacji), nazywa się uczeniem nienadzorowanym (ang. *unsupervised learning*)[80], czyli bez nauczyciela. Można je porównać do dominującej w okresie niemowlęcym nauki spontanicznej[81].

Niektóre algorytmy są w stanie przetwarzać częściowo oznakowane dane uczące, które najczęściej składają się z dużej ilości danych nieoznakowanych i stosunkowo niewielkiej liczby oznakowanych przykładów. Taką sytuację nazywamy uczeniem półnadzorowanym (ang. *semisupervised learning*)[82].

Twórcy systemów inteligentnych mają jednak trudności ze zrozumieniem tego, w jaki dokładnie sposób algorytmy uczenia maszynowego przetwarzają informacje w swo-

[77] Dzięki uczeniu głębokiemu badacze z Massachusetts Institute of Technology (MIT) byli w stanie skonstruować robota, który opanował grę w Jengę. Zob. J. Chu, *MIT robot combines vision and touch to learn the game of Jenga*, „MIT News", http://news.mit.edu/2019/robot-jenga-0130, 30.01.2019 r. (dostęp: 1.12.2024 r.).

[78] A. Géron, *Uczenie maszynowe z użyciem Scikit-Learn i TensorFlow*, tłum. K. Sawka, Gliwice 2018, s. 28.

[79] I. Goodfellow, Y. Bengio, A. Courville, *Deep learning*..., s. 102–103, 138–143.

[80] I. Goodfellow, Y. Bengio, A. Courville, *Deep learning*..., s. 102–103, 143–149.

[81] P. Psyllos, *Myśląca maszyna. Wstęp do sztucznej inteligencji*, cz. 2, 14.11.2024 r., https://petrospsyllos.com/pl/blog/598-myslaca-maszyna-wstep-do-sztucznej-inteligencji-cz-2-wideo/ (dostęp: 9.01.2025 r.).

[82] A. Géron, *Uczenie maszynowe*..., s. 33.

jej strukturze. Ta fundamentalna trudność, znana jako „problem czarnej skrzynki" (ang. *black box problem*), rodzi uzasadnione obawy co do możliwości wykorzystania programów uczących się w dziedzinach, gdzie kryteria dokonywania oceny mają istotne znaczenie (np. w praktyce wymiaru sprawiedliwości, medycynie). Powszechnie zgłaszane postulaty zwiększenia transparentności sposobu wnioskowania stały się dla badaczy impulsem do podjęcia starań o ustalenie, w jaki sposób algorytmy uczące się przetwarzają informacje źródłowe i jaką drogą dochodzą do takich, a nie innych wniosków[83].

1.4. Obecny stan sztucznej inteligencji

1.4.1. Wiele twarzy głębokiego uczenia

W ostatnich latach głębokie uczenie przekształciło się z czysto akademickiej dziedziny dla wtajemniczonych w siłę napędową Przemysłu 4.0, tworząc podstawę wielu komercyjnych zastosowań. Google LLC, jeden z pionierów głębokiego uczenia maszynowego, wykorzystuje uczenie głębokie w ogromnej liczbie swoich usług – od Tłumacza Google i Inbox Smart Reply przez Zdjęcia Google po Google Voice Search[84]. Jednak ani oferowane przez Google tłumaczenie w czasie rzeczywistym, ani autonomiczne formułowanie odpowiedzi na wiadomości e-mail, ani też zautomatyzowana organizacja zdjęć według przedstawionych na nich obiektów czy wyszukiwanie sterowane głosem nie wyczerpują bynajmniej szerokiego wachlarza powszechnie dostępnych usług internetowych, których podstawę tworzą metody uczenia głębokiego. Wyliczenie to nie daje również wyobrażenia o mnogości możliwych zastosowań tej technologii. Choć przedstawienie obszarów, w których jest ono wykorzystywane, przekracza ramy niniejszej monografii, wydaje się, że warto wymienić przynajmniej kilka z nich, by zobrazować przekrój możliwych zastosowań.

Metody uczenia maszynowego, w tym uczenia głębokiego, wykorzystywane są przez systemy rekomendacji, które analizują preferencje użytkowników i zdolne są nie tylko proponować filmy i albumy muzyczne z coraz większą trafnością odpowiadające ich gustom, lecz także zwiększać skuteczność działań reklamowych, dopasowując promo-

[83] W 2018 r. badacze z Massachusetts Institute of Technology (MIT) poddali dokładnej analizie sposób działania sieci służącej do klasyfikacji obiektów na obrazach. Zob. D. Bau, B. Zhou, A. Khosla, A. Oliva, A. Torralba, *Network Dissection: Quantifying Interpretability of Deep Visual Representations*, Massachusetts Institute of Technology, Computer Vision and Pattern Recognition (CVPR) 2017, http://netdissect.csail.mit.edu/ (dostęp: 1.12.2024 r.). Zwraca się przy tym uwagę, że ujawnienie sposobu, w jaki sztuczne sieci neuronowe przetwarzają informacje w swojej strukturze, nie leży w interesie korporacji wykorzystujących w zakresie swojej działalności algorytmy uczenia maszynowego. Zob. A. Burt, *The AI Transparency Paradox*, „Harvard Business Review", 3.12.2019 r., https://hbr.org/2019/12/the-ai-transparency-paradox (dostęp: 1.12.2024 r.).

[84] J. Le, *7 Machine Learning Applications at Google*, „Medium", 24.08.2016 r., https://medium.com/cracking-the-data-science-interview/7-machine-learning-applications-at-google-843d49d77bc8 (dostęp: 1.12.2024 r.).

wane produkty i usługi do konkretnego konsumenta. Programy uczące się są również nową siłą napędową rynków finansowych. Wykorzystywane są m.in. w procesie oceny zdolności kredytowej, do wychwytywania nowych trendów rynkowych z mediów społecznościowych czy poprawiania bezpieczeństwa biometrycznej bankowości[85]. Głębokie sieci neuronowe szkolone są także w zakresie diagnostyki medycznej. Skuteczność diagnostyczna niektórych systemów już dziś jest porównywalna z tą, którą osiągają lekarze specjaliści, a niekiedy nawet od niej lepsza. Uczenie maszynowe wykorzystywane jest również do analizowania ogromnych zbiorów danych generowanych przez różnorodne instrumenty naukowe, których przetworzenie przez badaczy metodami tradycyjnymi nie jest możliwe[86].

Nie ulega wątpliwości, że technologie sztucznej inteligencji, w szczególności te oparte na uczeniu głębokim, wywierają pozytywny wpływ na różnorodne dziedziny ludzkiej aktywności. Istnieją jednak także obszary, w których wykorzystanie sztucznej inteligencji budzi uzasadnione kontrowersje. Tak jest np. w dziedzinie przemysłu zbrojeniowego. Aktorzy współczesnych teatrów działań wojennych uzyskują wsparcie największych koncernów technologicznych w opracowywaniu technologii sztucznej inteligencji, które mogłyby zostać wykorzystane w działaniach zbrojnych[87].

1.4.2. Źródła sukcesu

Nie ulega wątpliwości, że systemy AI, których podstawę tworzą metody uczenia głębokiego, dynamicznie rozwijają się i są z powodzeniem wykorzystywane na wielu frontach. Skoro jednak, tak jak wcześniej wspomniano, pierwsze ANN powstały już w latach 50. ubiegłego wieku, można postawić pytanie: dlaczego uczenie głębokie dopiero niedawno osiągnęło dojrzałą postać? Dlaczego przez lata w dziedzinie AI prym wiodły podejścia oparte na systemach symbolicznych i dlaczego tyle trwało, nim oddały one palmę pierwszeństwa zaawansowanym metodom uczenia maszynowego? Można wskazać przynajmniej trzy powody takiego stanu rzeczy.

Po pierwsze, w ubiegłym stuleciu nie dysponowano odpowiednią ilością danych szkoleniowych. Siłą napędową współczesnych systemów opartych na metodach uczenia głębokiego są rosnące zbiory danych w postaci cyfrowej. To one nadają tempo ostatnim osiągnięciom w dziedzinie AI. Wzrost digitalizacji społeczeństwa spowodował, że ob-

[85] T.J. Sejnowski, *Deep Learning*..., s. 29.

[86] M. Young, *Machines Learning Astronomy: The new era of artificial intelligence & Big Data is changing how we do astronomy*, „Sky & Telescope", 22.11.2017 r., https://www.thefreelibrary.com/Machines+Learning+Astronomy%3A+The+new+era+of+artificial+intelligence+%26...-a0513926467 (dostęp: 1.12.2024 r.).

[87] Co do roli podmiotów sektora technologicznego w tworzeniu śmiercionośnej broni autonomicznej zob. raport *Don't be evil* przygotowany przez organizację PAX: F. Slijper, A. Beck, D. Kayser, M. Beenes, *Don't be evil*, 19.08.2019 r., PAX, https://www.paxforpeace.nl/publications/all-publications/dont-be-evil (dostęp: 1.12.2024 r.).

serwujemy dziś napływ ogromnych ilości danych cyfrowych, pochodzących z różnorodnych źródeł informacyjnych[88]. Kopalnią danych są popularne portale społecznościowe, komunikatory i rozmaite serwisy internetowe.

Według danych z 2023 r., w każdej minucie wysyłanych jest ponad 240 mln e-maili, a liczba operacji wyszukania w Google sięga 6,3 mln[89]. Ogromne ilości danych generowane są każdego dnia przez wszystkie urządzenia podłączone do sieci, niezależnie od tego, czy jest to pojazd, sprzęt AGD, czy urządzenie mobilne takie jak smartfon albo zegarek typu smartwatch. Stały przyrost danych w postaci cyfrowej jest oczywistą konsekwencją rozwoju technologicznego. Tempo tego przyrostu jest jednak oszałamiające. W 1992 r. na świecie powstawało 100 gigabajtów (10^9 bajtów) danych dziennie, w 1997 r. do wygenerowania tej samej ilości danych wystarczyła godzina, a w 2002 r. – sekunda[90]. W 2023 r. każdego dnia globalnie generowanych było niemal 329 eksabajtów (10^{18} bajtów) danych, natomiast szacuje się, że w 2025 r. ilość danych generowanych dziennie osiągnie 181 zattabajtów (10^{21} bajtów) danych[91]. Dane wykorzystywane w procesie uczenia algorytmów przybierają przy tym niezliczoną różnorodność form[92]. Niemal wszystko to, co może zostać zgromadzone i przedstawione w formie cyfrowej – aktywność na Facebooku, transakcje dokonywane za pomocą karty płatniczej, rozmowy telefoniczne, trasy sezonowych migracji żółwi morskich, dane statystyczne dotyczące wyroków sądowych, karty informacyjne z leczenia szpitalnego – stanowi potencjalne dane szkoleniowe dla algorytmów uczących się.

Po drugie, technologie komputerowe w ubiegłym stuleciu nie były jeszcze wystarczająco wydajne. Wzrost wykładniczy mocy obliczeniowej kolejnych generacji komputerów, możliwy dzięki postępującej miniaturyzacji tranzystorów, trwa nieprzerwanie od lat 60. ubiegłego wieku. Skonstruowany w tamtym czasie CDC 6600, uważany za pierwszy superkomputer, mógł poszczycić się mocą obliczeniową rzędu 3 MFLOPS (10^6 FLOPS), co oznacza, że był w stanie wykonać 3 mln operacji zmiennoprzecinkowych na sekundę. Dla porównania smartfon Samsung Galaxy S6 osiąga moc 34,8 GFLOPS (10^9 FLOPS). Moc obliczeniowa dzisiejszych superkomputerów jest nieporównywalnie większa – IBM Summit wykonuje 200 biliardów operacji na sekundę, osiągając moc 200 PFLOPS (10^{15} FLOPS)[93]. Z prędkością odpowiadającą tempu wzrostu mocy obliczeniowej zwiększa się również pamięć komputerowa. Pamięć kasowalna (odpowiednik dzisiejszego RAM-u) opracowanego w 1966 r. komputera, który znalazł się na pokładzie Apollo 11

[88] M. Tabakow, J. Korczak, B. Franczyk, *Big Data – definicje, wyzwania i technologie informatyczne*, „Informatyka Ekonomiczna" 2014/1(31), s. 138.

[89] E. Griifith, *What Happens in 60 Seconds on the Internet*?, „PCMag", 14.12.2023 r., https://www.pcmag.com/news/what-happens-in-60-seconds-of-global-internet-activity (dostęp: 1.12.2024 r.).

[90] Zob. https://www.forbes.pl/autorzy/samsung-mm-134000 (dostęp: 1.12.2024 r.).

[91] F. Duarte, *Amount of Data Created Daily (2024)*, „Exploding Topics", 13.06.2024 r., https://explodingtopics.com/blog/data-generated-per-day#how-much (dostęp: 1.12.2024 r.).

[92] J. Kaplan, *Sztuczna inteligencja*..., s. 48.

[93] Zob. https://www.ibm.com/thought-leadership/summit-supercomputer/ (dostęp: 1.12.2024 r.).

(ang. *Apollo Guidance Computer*), miała pojemność 4 kB (10^3 bajtów)[94], czyli dwa miliony razy mniej niż współczesne laptopy i dwa biliony pięćset miliardów razy mniej niż superkomputer Summit. Pierwszy twardy dysk – stworzony w 1956 r. IBM 350 – miał pojemność 5 MB (10^6 bajtów). Pierwszy komercyjny komputer IBM 305 RAMAC, wyposażony w dysk typu 305, kosztował 160 000 USD (dziś ok. 1 241 000 USD). W 1981 r. dysk twardy Apple o tej samej pojemności kosztował 3 500 USD (dziś ok. 9 840 USD). Obecnie nośnik o pojemności 5 MB jest niemal bezwartościowy – dyski o milionkrotnie większej pojemności można kupić za ok. 100 USD.

Po trzecie, dostępność bardziej zaawansowanej infrastruktury komputerowej pozwala na opracowywanie znacznie większych modeli ANN. Rozrastanie się sieci następuje również w tempie wykładniczym. Do niedawna sieci neuronowe tworzyło relatywnie niewiele sztucznych neuronów. Jak podnoszą I. Goodfellow, Y. Bengio i A. Courville, nie powinno dziwić, że ANN lat 80. i 90., mające mniej neuronów niż – odpowiednio – dżdżownice i pijawki, nie były zdolne do wykonywania skomplikowanych zadań. Choć, jak zauważają, „nawet dzisiejsze sieci, dość duże z obliczeniowego punktu widzenia, mają system nerwowy mniejszy niż względnie prymitywne zwierzęta, takie jak żaba”[95], utrzymująca się tendencja wzrostu rozmiaru sztucznych sieci neuronowych pozwala przypuszczać, że osiągną one liczbę neuronów odpowiadającą liczbie organicznych neuronów w ludzkim mózgu w ciągu kolejnych 30 lat.

1.4.3. Ogólna i wąska sztuczna inteligencja

Wszystkie dotychczasowe osiągnięcia w dziedzinie AI należą do kategorii tzw. wąskiej sztucznej inteligencji (ang. *Narrow Artificial Intelligence*). Oznacza to, że zastosowanie istniejących systemów ogranicza się do konkretnych, wąskich dziedzin. Realizują one pewne specyficzne cele, takie jak granie w gry czy rozwiązywanie zadań arytmetycznych, osiągając przy tym poziom porównywalny do ludzkiego lub od niego wyższy.

Poza zasięgiem badaczy wciąż pozostaje tzw. ogólna sztuczna inteligencja (ang. *General Artificial Intelligence*), którą M. Tegmark określa mianem Świętego Graala tej dziedziny. Ogólna sztuczna inteligencja zdolna jest do rozwiązania każdego problemu, podobnie jak człowiek, który jeśli tylko będzie trenował wystarczająco długo, może zrealizować dowolny cel: zagrać w każdą grę, nauczyć się każdego języka, zostać specjalistą w każdej dziedzinie (od fizyki doświadczalnej po kardiochirurgię).

[94] P. Urbaniak, *Apollo Guidance Computer. Komputer, który zabrał człowieka na Księżyc*, „Komputer Świat”, 20.07.2019 r., https://www.dobreprogramy.pl/apollo-guidance-computer-komputer-ktory-zabral-czlowieka-na-ksiezyc,6628573828814977a (dostęp: 1.12.2024 r.).

[95] I. Goodfellow, Y. Bengio, A. Courville, *Deep learning*..., s. 22–23.

1.5. Filozofia sztucznej inteligencji

1.5.1. Wielkie nadzieje i obawy

Z dotychczasowych rozważań wynika, że znane systemy AI są niczym innym jak wyrafinowanym pokazem zdolności inżynieryjnych. Bez względu na to, jak zdumiewające umiejętności demonstrują programy komputerowe, które określamy dziś mianem inteligentnych, pozostają one jedynie zaawansowanymi technologicznie narzędziami i kolejnym krokiem w postępie automatyzacji.

Dlaczego więc AI wzbudza tak silne kontrowersje, podczas gdy osiągnięcia innych gałęzi inżynierii nie stają w ogniu tak silnej krytyki? Odpowiedź nasuwa się intuicyjnie: te ostatnie nie jawią się nam jako zagrożenie dla transcendentnej natury istoty ludzkiej. Obawy, że osiągnięcia w dziedzinie AI mogą podważyć znany nam porządek świata, w którym to człowiek zajmuje pozycję nadrzędną, są w pewnej mierze skutkiem przesadnie optymistycznych prognoz dotyczących tempa rozwoju tej technologii i jej możliwości[96]. J. Kaplan dostrzega, że nasze rozumienie AI zaćmiewają aspiracyjne związki między inteligencją maszynową i ludzką. Aby wyjaśnić tę myśl, proponuje on wyobrazić sobie kontrowersje, jakie mogłyby zapanować w lotnictwie, gdyby samoloty z napędem silnikowym od początku nazywano „sztucznymi ptakami".

> „Gdyby to mylne ujęcie się utrwaliło, mogłyby się odbywać konferencje z udziałem ekspertów i znanych postaci zastanawiających się, co się stanie, kiedy samoloty nauczą się wić gniazda, rozwijać umiejętność projektowania i budowania swojego potomstwa, poszukiwać paliwa, by nakarmić swoje młode i tak dalej"[97].

Niezależnie od tego, jak absurdalna wydawać się może ta wizja, J. Kaplan dostrzega w niej analogię do aktualnych obaw dotyczących AI. Podczas gdy zasadniczo nic poza tym, co sam nazywa „najdzikszymi spekulacjami", nie daje podstaw do obawiania się detronizacji człowieka przez komputery przyszłości, wyobrażenie o nieuniknionym „buncie robotów", karmiące się literaturą i filmami science-fiction, jest w naszych umysłach żywe. Można zastanawiać się, czy sztuczna inteligencja rodziłaby porównywalne obawy, gdyby J. McCarthy wybrał dla jej określenia „bardziej przyziemny termin, który nie sugeruje wyzwania dla ludzkiej dominacji czy poznania, taki jak «przetwarzanie symboliczne», albo «obliczenia analityczne»"[98].

---

[96] S. Armstrong, K. Sotala, S.S. Oh' Eigeartaigh, *The errors, insights and lessons of famous AI predictions – and what they mean for the future*, „Journal of Experimental & Theoretical Artificial Intelligence" 2014/3(26), s. 329.

[97] J. Kaplan, *Sztuczna inteligencja*..., s. 33–34.

[98] J. Kaplan, *Sztuczna inteligencja*..., s. 34.

1.5.2. Czy komputery potrafią myśleć? Test Turinga i „chiński pokój"

Pierwsi badacze sztucznej inteligencji zastanawiali się, kiedy system skonstruowany przez człowieka można nazwać inteligentnym. Pytanie to po dziś dzień pozostaje jednym z kluczowych problemów filozofii AI. Refleksja nad tym, co czyni człowieka inteligentnym, doprowadziła do wykształcenia się dwóch odmiennych poglądów na istotę inteligencji maszynowej. Pierwszy z nich zakłada, że komputery są albo w przyszłości będą umysłami, a przynajmniej posiądą wszystkie cechy ludzkiego umysłu. Taki rodzaj sztucznej inteligencji określa się mianem silnej (ang. *Strong Artificial Intelligence*). Zgodnie z drugim poglądem komputery nie są i nigdy nie będą umysłami, a jedynie symulują właściwości umysłu, które uznajemy za przejaw inteligencji. Jest to tzw. słaba sztuczna inteligencja (ang. *Weak Artificial Intelligence*)[99]. Powyższe rozróżnienie, jak zwięźle ujmuje to J. Kaplan, dotyczy pytania „czy maszyny mogą być naprawdę inteligentne, czy tylko zdolne do działania «jak gdyby» były inteligentne?"[100].

Jeden z ojców współczesnej informatyki, wybitny angielski matematyk A. Turing, zaproponował eksperyment myślowy oparty na regułach gry imitacyjnej[101], który dał podwaliny popularnej dziś metodzie wykorzystywanej do oceny zdolności AI, znanej powszechnie jako Test Turinga. Jego idea polega na tym, że człowiek-sędzia prowadzi za pomocą urządzenia uniemożliwiającego zidentyfikowanie rozmówcy dialog z innym człowiekiem oraz z komputerem równocześnie. Zadanie sędziego polega na rozpoznaniu interlokutorów, to jest stwierdzeniu, które wypowiedzi pochodzą od człowieka, a które od komputera[102]. Jeśli sędzia nie potrafi odróżnić wypowiedzi człowieka od tych pochodzących od komputera, to możemy powiedzieć, że komputer „myśli". Choć Test Turinga został pierwotnie pomyślany dla wąskiego obszaru AI, jakim jest przetwarzanie języka naturalnego, mógłby on z powodzeniem zostać wykorzystany także do oceny innych obszarów zastosowania sztucznej inteligencji.

Początkowe sukcesy w zakresie konstruowania systemów symbolicznej sztucznej inteligencji pchnęły A. Newella i H.A. Simona do sformułowania w 1976 r. hipotezy tzw. fizycznego systemu symbolicznego (ang. *physical symbol system*), zgodnie z którą „fizyczny system symboliczny ma konieczne i wystarczające środki do wykonywania inteligentnych działań"[103]. Zarysowując tę hipotezę, wpisującą się w nurt silnego podejścia w sztucznej inteligencji, A. Newell i H.A. Simon podnieśli, że każdy system, który „przejawia ogólną inteligencję, okaże się po analizie fizycznym systemem sym-

[99] J. Kaplan, *Sztuczna inteligencja*..., s. 91.

[100] J. Kaplan, *Sztuczna inteligencja*..., s. 91.

[101] A.M. Turing, *Computing Machinery and Intelligence,* „Mind: A Quarterly Review of Psychology and Philosophy" 1950/59(236), s. 433 i n.

[102] M. Flasiński, *Wstęp*..., s. 3–4.

[103] A. Newell, H.A. Simon, *Computer Science as Empirical Inquiry: Symbols and Search*, „Communications of the ACM" 1976/3(19), s. 116.

bolicznym", a „każdy fizyczny system symboliczny wystarczających rozmiarów może być zorganizowany tak, by przejawiać ogólną inteligencję"[104]. Jeśli rozumieć myślenie jako zdolność do manipulowania symbolami w celu przeprowadzenia rozumowania, czyli wyciągania poprawnych wniosków z początkowych założeń, to stwierdzenie, że programy przetwarzające symbole myślą (niezależnie od tego, jak trywialne by one nie były), zasługuje na aprobatę.

W celu odrzucenia hipotezy fizycznego systemu symbolicznego filozof J.R. Searle, jeden z krytyków silnej sztucznej inteligencji, zaproponował eksperyment myślowy zwany „chińskim pokojem" (ang. *Chinese Room*). W pewien sposób ów eksperyment przypomina opisany już Test Turinga. Załóżmy, że w pokoju zostaje zamknięty człowiek-tłumacz, który nie zna języka chińskiego. W pomieszczeniu znajdują się jednak książki zawierające sformalizowane instrukcje, umożliwiające odczytywanie chińskich symboli i generowanie nowych zapisów. Tłumacz otrzymuje pytania zapisane przy pomocy chińskich symboli, a następnie udziela na nie odpowiedzi, posługując się instrukcjami zawartymi w książkach. Następnie odpowiedzi przekazywane są na zewnątrz pokoju. Przyjmijmy teraz, że osoba, która zadaje pytania i zna język chiński, nie jest w stanie odróżnić odpowiedzi tłumacza od odpowiedzi osoby, która naprawdę potrafi się tym językiem posługiwać. Czy możemy powiedzieć, że tłumacz zna chiński? Oczywiście, że nie. Według J.R. Searle'a tłumacz odgrywa w tym eksperymencie rolę odpowiednio zaprogramowanego komputera, który choć jest w stanie udzielać poprawnych odpowiedzi, w rzeczywistości nie rozumie, co robi, tak jak wymagałaby tego silna sztuczna inteligencja[105].

Manipulacja symbolami nie jest według J.R. Searle'a myśleniem. Można by jednak zastanowić się, jaka jest różnica między kłębiącymi się w naszych mózgach myślami i danymi bitowymi przetwarzanymi przez komputer. Otóż *prima facie* żadna. Konstatacja ta prowadzi jednak do pewnego paradoksu, który J. Kaplan ujmuje następująco:

> „[J]eśli jesteśmy przekonani, że nasze mózgi nie są niczym innym jak układami, które manipulują symbolami i składają się z materiału biologicznego, to w naturalny sposób jesteśmy zmuszeni do wyciągnięcia wniosku, że nasz mózg sam przez się nie może myśleć. [...] Albo jedno, albo drugie: jeśli manipulowanie symbolami jest podstawą inteligencji, to albo zarówno ludzie, jak i maszyny mogą myśleć [...], albo ani ludzie, ani maszyny myśleć nie mogą"[106].

[104] A. Newell, H.A. Simon, *Computer Science*..., s. 116.

[105] Informacje w akapicie za: J.R. Searle, *Minds, brains, and programs*, „Behavioral and Brain Sciences" 1980/3(3), s. 417 i n. Należy dostrzec również argumenty wysuwane przeciwko „chińskiemu pokojowi". Zob. D. Cole, *The Chinese Room Argument*, „Stanford Encyclopedia of Philosophy", 23.10.2024 r., https://plato.stanford.edu/entries/chinese-room/ (dostęp: 9.01.2025 r.).

[106] J. Kaplan, *Sztuczna inteligencja*..., s. 97–98.