



OLIWIA KRÓLIKIEWICZ

Uniwersytet Komisji Edukacji Narodowej w Krakowie

 <https://orcid.org/0009-0002-8914-7203>

Sztuczna inteligencja w ujęciu odpowiedzialności karnej – luka w odpowiedzialności czy możliwość jej przypisania?

Artificial intelligence in terms of criminal liability – a gap in liability or the possibility of assigning it?

Abstract: The article provides an examination of the rationale for adopting criminal responsibility for artificial intelligence systems. It identifies the basic concepts associated with artificial intelligence and goes on to explore its relationship to morality and ethics. The different approaches to adopting a framework of criminal responsibility for artificial intelligence are discussed in this article, and this is achieved by means of an analysis of the theoretical foundations of criminal responsibility in the continental and common law systems. The article examines the application of familiar criminal science concepts, such as guilt, consciousness and the act, to artificial intelligence systems, and it answers the question of mitigating the liability gap by presenting alternative approaches to the problem for researchers within the Continental and Anglo-Saxon systems. In the concluding section, the article comments on the interpenetration of civil law (and its institutions) with criminal law.

Keywords: artificial intelligence (AI), responsibility, ethics, morality, criminal responsibility, accountability gap, fault

Im większa niezależność maszyny,
tym większa potrzeba zakreślenia jej granic moralnych.

Rosalind Picard¹

Nietrudno sobie wyobrazić, jak technologia przechytrza rynki finansowe i badaczy, manipuluje ludzkimi przywódcami oraz tworzy broń, której natury nawet nie rozumiemy. Podczas gdy krótkoterminowy wpływ sztucznej inteligencji zależy od tego, kto ją w danym momencie kontroluje, długoterminowy wpływ zależy od tego, czy w ogóle da się ją kontrolować.

Stephen Hawking, Stuart Russell, Max Tegmark, Frank Wilczek²

1. Wprowadzenie

Jako ludzkość stoimy przed doniosłym pytaniem o *ratio* kontynuowania prac nad rozwojem innowacyjnych technologii samouczących się i innych proliferacji nieludzkich bytów³ opartych na sztucznej inteligencji, zdolnych do projektowania i stosowania systemów niepotrzebujących do swych prac angażowania człowieka. Jednocześnie możemy podjąć zgoła odmienną decyzję o powstrzymaniu się od wprowadzania systemów uczenia do społeczeństwa i zachowaniu pełnej odpowiedzialności człowieka⁴. Zgodnie z teorią instrumentalną kwestia odpowiedzialności za działania nowych technologii spoczywa na człowieku, gdyż nowe technologie traktuje on jako zwykłe narzędzie⁵ lub przedmiot ludzkiego działania⁶. Sprawa jednak komplikuje się w przypadku najnowocześniejszych technologii i robotów społecznych, które dzięki przetwarzaniu języka naturalnego

¹ R. Picard: *Teaching Robots Right from Wrong*. "The Economist", 05.07.2017. <https://www.economist.com/1843/2017/07/05/teaching-robots-right-from-wrong> [dostęp: 09.07.2024]; jeśli nie wskazano inaczej, cytaty podano w tłum. własnym.

² S. Hawking: *Transcendence Looks at the Implication of Artificial Intelligence – But Are We Taking AI Seriously Enough?*. "The Independent", 01.05.2014. <https://www.independent.co.uk/news/science/stephen-hawking-transcendence-looks-at-the-implications-of-artificial-intelligence-but-are-we-taking-ai-seriously-enough-9313474.html> [dostęp: 09.07.2024].

³ M. Coeckelbergh: *Virtual Moral Agency, Virtual Moral Responsibility: On the Moral Significance of the Appearance, Perception, and Performance of Artificial Agents*. "AI and Society" 2009, vol. 24, s. 181–189. <https://doi.org/10.1007/s00146-009-0208-3>.

⁴ F. Santoni de Sio, G. Mecacci: *Four Responsibility Gaps with Artificial Intelligence: Why they Matter and How to Address them*. "Philosophy & Technology" 2021, vol. 34, s. 1057–1084. <https://doi.org/10.1007/s13347-021-00450-x>.

⁵ Za traktowaniem robota jako narzędzie w rękach człowieka: D.J. Gunkel: *Robot Rights*. The MIT Press, Cambridge 2018, s. 72.

⁶ S.M. Solaiman: *Legal Personality of Robots, Corporations, Idols and Chimpanzees: A Quest for Legitimacy*. "Artificial Intelligence and Law" 2017, no. 25 (2), s. 177. <https://doi.org/10.1007/s10506-016-9192-3>.

podważają tradycyjne podejście do przedmiotu⁷. Do tej pory sprawczość moralna była postrzegana wyłącznie jako cecha ludzka⁸. Czy istnieje możliwość włączenia sztucznych czynników do dyskursu moralnego?

Sztuczna inteligencja (SI, ang. *artificial intelligence* – AI) jest działem informatyki zajmującym się konstruowaniem maszyn i algorytmów, których działanie wykazuje znamiona inteligencji. Przedmiotem działania AI jest badanie i określanie reguł, które rządzą inteligentnymi zachowaniami człowieka, oraz wykorzystywanie ich w algorytmach i programach komputerowych⁹. Obecnie rozwój sztucznej inteligencji wpływa na optymalizację łańcucha dostaw i planowanie zasobów w celu redukcji emisji dwutlenku węgla, a także tworzone są aplikacje mające duże znaczenie dla ludzkości, związane z odkrywaniem leków, znajdowaniem lepszych metod leczenia pacjentów, nanoinżynierią, diagnostyką medyczną, prognozowaniem systemów pogodowych¹⁰ czy optymalizacją alokacji ograniczonych zasobów¹¹.

Według Herberta L.A. Harta człowiek jest prawnie odpowiedzialny za czyn lub szkodę po stwierdzeniu, że jego związek z czynem lub szkodą jest zgodnie z prawem wystarczający do pociągnięcia go do odpowiedzialności. Ponieważ osoba jest prawnie odpowiedzialna za jakieś działanie, może również zostać za nie ukarana¹². Rozwój nowych technologii opartych na sztucznej inteligencji rodzi wiele pytań dotyczących odpowiedzialności za przestępstwa związane z jej działaniem. Udoskonalane technologie oznaczają się coraz większą autonomią i złożonością, a wpływ człowieka staje się ograniczony. Użyteczność rozwiązań bazujących na AI przewyższa potencjalne ryzyko wystąpienia szkód. Celem artykułu jest odpowiedzenie na pytanie, czy w odniesieniu do sztucznej inteligencji możemy mówić o jej odpowiedzialności karnej, czy też występuje tzw. luka w odpowiedzialności. Idei tej służyć ma zbadanie teoretycznych podstaw przypisania odpowiedzialności karnej z uwzględnieniem *common law* i prawa karnego w systemie kontynentalnym.

⁷ D.J. Gunkel: *How to Survive a Robot Invasion. Rights, Responsibility, and AI*. Routledge, New York 2020, s. 70.

⁸ M. Coeckelbergh: *Virtual Moral Agency...*, s. 181.

⁹ M. Czochra, D. Bar: *Śmierć pacjenta wywołana zastosowaniem sztucznej inteligencji w technologiach medycznych – analiza prawnokarna*. „Studia Prawnicze. Rozprawy i Materiały” 2019, nr 2 (25), s. 67.

¹⁰ D.J. Baker, P.H. Robinson: *Artificial Intelligence and the Law: Cybercrime and Criminal Liability*. Routledge, New York 2021, s. 221.

¹¹ *AI in eHealth. Human Autonomy, Data Governance and Privacy in Healthcare*. Eds. M.C. Compagnucci i in. Cambridge University Press, Cambridge 2022, s. 300.

¹² H.L.A. Hart: *Punishment and Responsibility: Essays in the Philosophy of Law*. 2nd ed. Oxford University Press, Oxford 2008.

2. Autonomiczność decyzji systemów AI w aspekcie prawa karnego

Czy istnieje możliwość włączenia sztucznych czynników do dyskursu moralnego? Jednym z przymiotów sztucznej inteligencji jest autonomiczność działania. Autonomia sztucznej inteligencji odnosi się do zdolności systemu AI do podejmowania określonych zachowań bez wyraźnych poleceń zewnętrznych oraz niezależności od innych systemów AI¹³.

Szczególnie nieprzewidywalne jest użycie autonomicznych pojazdów i autonomicznych systemów uzbrojenia (ang. *autonomous weapon systems* – AWS)¹⁴, które podejmują „własne”¹⁵ decyzje, niezależne od ludzkich operatorów. Niesie to ze sobą pytania i jeszcze bardziej skomplikowane odpowiedzi na temat odpowiedzialności za ich niewłaściwe bądź niecelowe zastosowanie, będące realnym zagrożeniem dla życia i zdrowia ludzi. Potrącenia pieszych, zakup nielegalnych substancji czy powodowanie krachu na rynkach finansowych – to wszystko może stać się udziałem systemu AI.

Istnieje coraz większa klasa działań maszyn, w przypadku których tradycyjne sposoby przypisywania odpowiedzialności nie są zgodne z poczuciem sprawiedliwości czy moralnymi ramami społeczeństwa, ponieważ nikt nie ma wystarczającej kontroli nad działaniami maszyn, by móc przejąć odpowiedzialność za nie¹⁶. Już od lat 70. w Stanach Zjednoczonych podnoszono problem etyki komputerowej, a wraz z nią kwestie autonomicznych czynników poznawczych i moralnych¹⁷. AI jest technologią rewolucyjną, a związane z nią konsekwencje w zakresie przypisania winy są doniosłe.

¹³ V. Dignum: *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Springer, Cham 2019, s. 18.

¹⁴ W przypadku broni autonomicznej podnosi się argument, że podważa ona elementarne zobowiązanie do odpowiedzialności, szczególnie wtedy, gdy nie ma możliwości moralnego przypisania jej jednostce ludzkiej za zdarzenie spowodowane przez AWS. Powstaje zatem luka w odpowiedzialności, której nie można przypisać ani producentowi, ani programiście, ani z wiadomych względów – samej autonomicznej broni. *The Cambridge Handbook of Responsible Artificial Intelligence, Interdisciplinary Perspectives*. Eds. S. Voenekey i in. Cambridge University Press, Cambridge 2022, s. 478.

¹⁵ *Ethics of Socially Disruptive Technologies: An Introduction*. Eds. I. van de Poel i in. Cambridge University Press, Cambridge 2023, s. 21.

¹⁶ D.J. Gunkel: *How to Survive...*, s. 43.

¹⁷ S. Quattrocolo: *Artificial Intelligence, Computational Modelling and Criminal Proceedings: A Framework for A European Legal Discussion*. Springer, Cham 2020, s. 122.

3. Luka w odpowiedzialności

Obecnie problem luki w odpowiedzialności uważany jest za problem prawny¹⁸. Luka w odpowiedzialności (inaczej: zerwanie łańcucha odpowiedzialności, ang. *responsibility gap*)¹⁹ oznacza niemożność wskazania podmiotu odpowiedzialnego²⁰, który powinien zrekompensować szkodę, w sytuacjach, w które zaangażowany jest także system sztucznej inteligencji²¹. Tworzy to trzy rodzaje ryzyk: ryzyko autonomii (proces samouczenia się sztucznej inteligencji sprawia, że AI zmienia swój program, staje się niezależna i nieprzewidywalna), ryzyko powiązania (interakcja pomiędzy ludźmi a systemami AI jest złożona i utrudnia zrozumienie dokładnego podziału ról i zadań między nimi, co może prowadzić do pojawienia się efektu *black box effect*) oraz ryzyko sieciowe (uwzględnia złożone interakcje między systemami sztucznej inteligencji, tworząc zespół wielu aktywnych i polimorficznych inteligentnych agentów)²². Wykorzystanie sztucznej inteligencji, pomimo jej niekwestionowanej mocy, może także nieść ryzyko powstania szkody. Wraz z rozwojem sztucznej inteligencji przestępstwa mogą być popełniane przy użyciu nowych technologii albo przez nie same.

Droga do osiągnięcia sprawiedliwego podziału ryzyka i winy w kontekście nowych technologii pozostaje wciąż otwarta²³. „AI może pociągać za sobą nowe formy odpowiedzialności karnej, nowe luki w prawie karnym lub nowe wyzwania dla praw oskarżonych o przestępstwa do

¹⁸ *Regulating Artificial Intelligence in Industry*. Ed. D.M. Bielicki. Routledge, New York 2021, s. 204.

¹⁹ Luka w odpowiedzialności oznacza, że związek pomiędzy działaniem systemu sztucznej inteligencji a wolą i świadomością człowieka w kontekście autonomicznych systemów AI całkowicie zanika. R. Rejmank: *Autonomiczność systemów sztucznej inteligencji jako wyzwanie dla prawa karnego*. „Roczniki Nauk Prawnych” 2021, t. 31, nr 3, s. 108.

²⁰ A. Beckers, G. Teubner: *Three Liability Regimes for Artificial Intelligence*. Hart Publishing, Oxford 2021, s. 97.

²¹ Konieczne staje się udzielenie odpowiedzi na pytanie, czy doszło do przestępstwa, oraz ustalenie kręgu osób odpowiedzialnych. „Złożony proces innowacyjnych podzespołów wzajemnie powiązanych i składających się na urządzenie medyczne jest bowiem przedsięwzięciem, w którym uczestniczy kilkudziesięciu lub nawet kilkuset specjalistów z różnych dziedzin nauki. Tymczasem przypisanie winy w prawie karnym musi wiązać się z jej indywidualizacją i konkretnym określeniem podmiotu odpowiedzialnego, co w ujęciu praktycznym mogłoby uczynić proces pociągnięcia do odpowiedzialności karnej określenia osoby wyjątkowo trudnym”. M. Czochoła, D. Bar: *Śmierć pacjenta...*, s. 77.

²² K. Prifti, E. Stamhuis, K. Heine: *Digging into the Accountability Gap: Operator's Civil Liability in Healthcare AI-systems*. W: *Law and Artificial Intelligence: Regulating AI and Applying AI in Legal Practice*. Eds. B. Custers, E. Fosch-Villaronga. Springer, Berlin 2022, s. 280.

²³ F. Santoni de Sio, G. Mecacci: *Four Responsibility Gaps...*, s. 1072.

sprawiedliwego procesu”²⁴. Systemom sztucznej inteligencji brakuje zdolności refleksyjnych i świadomości własnych działań, które dziś powszechnie uważamy za niezbędne do pociągnięcia do odpowiedzialności²⁵.

Zagadnienie luki w odpowiedzialności odnoszącej się do nauczania automatycznego podjął w nauce Andreas Matthias, który zauważył problem rosnących zdolności uczenia się inteligentnych maszyn i ich wpływu na decyzyjność oraz kontrolę człowieka²⁶. Luki w odpowiedzialności stanowią doniosłe wyzwanie dla współczesnego prawa, moralności i etyki. Powstaniu i pogłębianiu się luki w odpowiedzialności sprzyjają: nieprzewidywalność, złożoność, nieprzejrzystość (tzw. *black box effect*), interaktywność zachowania sztucznej inteligencji oraz brak interpretowalności podejmowania decyzji wspomaganych maszynowo.

W piśmiennictwie dominują trzy podejścia. Pierwsze, reprezentowane m.in. przez A. Matthiasa, podejście fatalistyczne, odnosi się do uznania nowych technologii opartych na sztucznej inteligencji za wydarzenie bezprecedensowe, nowe, które jest wyzwaniem dla odpowiedzialności moralnej (winy) i które w zasadzie jest prawnie nierozwiązywalne²⁷.

Na drugim krańcu debaty mieści się pogląd stanowiący o tym, że luka w odpowiedzialności za działania automatycznych maszyn napędzanych sztuczną inteligencją nie jest problemem nowym, a do jego rozwiązania wystarczą sprawdzone i istniejące już moralno-prawne metody. Stanowisko to głosi pochwałę wprowadzania nowych technologii, nawet jeśli skutkiem tego będzie erozja ludzkiej odpowiedzialności moralnej. Popierający ten koncept skłaniają się ku opinii, wedle której nie zawsze i nie za wszelką cenę ustalenie winy będzie możliwe, a zatem powinien powstać pewien margines, w zakresie którego niemożność ustalenia winy będzie nieunikniona²⁸.

Trzecim modelem, solucjonistycznym, jest koncept dążący do rozwiązania problemu luki w odpowiedzialności poprzez zaproponowanie nowych możliwości technicznych i prawnych. Takie próby podejmowane są już teraz przez Unię Europejską w kontekście odpowiedzialności cywilnej²⁹. Ich rozwinięciem jest propozycja wdrożenia bardziej kompleksowych

²⁴ S. Quattrocolo: *Artificial Intelligence...*, s. 7.

²⁵ *Ethics of Socially Disruptive Technologies...*, s. 21.

²⁶ A. Matthias: *The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata*. "Ethics and Information Technology" 2004, no. 6 (3), s. 175.

²⁷ Ibidem.

²⁸ F. Santoni de Sio, G. Mecacci: *Four Responsibility Gaps...*, s. 1071.

²⁹ Chodzi o: AIA – Artificial Intelligence Act (Akt w sprawie sztucznej inteligencji) z dnia 13 marca 2024 r. – pierwsze w ujęciu globalnym ramy prawne dla sztucznej inteligencji; Revised Product Liability Directive – ewaluacja PLD, przyjęta przez Parlament Europejski dnia 13 marca 2024 r.; AILD (Dyrektywa PE ws. odpowiedzialności za sztuczną inteligencję – projekt). H. van Kolschooten: *EU Regulation of Artificial Intelligence*,

rozwiązań, których trzonem byłoby utrzymanie znaczącej kontroli człowieka nad sztuczną inteligencją, co *per se* wyeliminowałoby lukę w odpowiedzialności. Sprzyjać temu ma system śledzenia (ang. *tracking*), kontroli i nadzoru i jego projektowanie tak, by sposób działania był przejrzysty, poszczególne decyzje identyfikowalne i zrozumiałe oraz przynajmniej jeden czynnik ludzki wyodrębnialny³⁰. O ile w wojskowej automatyzacji ma to swoje uzasadnienie, to całkowita kontrola człowieka w innych sektorach może spowalniać innowacje.

4. Gabriel Hallevy i trzy modele odpowiedzialności

Szybki rozwój technologii sprawia, że każda norma prawna może szybko stać się nieaktualna. Istnieje wyraźna potrzeba zdefiniowania nowych kategorii przestępstw. Ze względu na transgraniczny charakter nowych technologii określenie ogólnych zasad nimi rządzących może okazać się skomplikowane³¹. Gabriel Hallevy w swych pracach wyodrębnił trzy modele odpowiedzialności: model odpowiedzialności za przestępstwa popełnione przez inny podmiot, model odpowiedzialności za przestępstwo popełnione z naturalnym prawdopodobieństwem oraz model bezpośredniej odpowiedzialności systemu sztucznej inteligencji.

Pierwszy to model, wedle którego za dane przestępstwo odpowiedzialność ponosi inny podmiot. Może być nim programista AI (np. zatrudniony programista opracowuje program, zadaniem którego jest wzniesienie ognia w fabryce), jak również użytkownik tego systemu (np. użytkownik posługuje się robotem-asystentem, któremu zleca zadania wykraczające poza powszechnie uznawane standardy zachowania). W obydwu przypadkach popełnienia przestępstwa fizycznie dopuścił się system sztucznej inteligencji. Zarówno u użytkownika, jak i u programisty nie wystąpiła ani fizyczna, ani psychiczna manifestacja elementu konkretnego przestępstwa. W tym modelu zachowanie programisty i użytkownika zrównuje się z zachowaniem samego systemu sztucznej inteligencji. Uzasadnieniem dla przyjęcia takiego rozwiązania jest posłużenie się AI jako narzędziem

Challenges for Patients' Rights. "Amsterdam Law School Legal Studies Research Paper, Kluwer Law International" 2021, vol. 59, s. 81.

³⁰ Prawo karne jest niewystarczające w przypadkach, gdy sztuczna inteligencja powoduje szkody, w których nie sposób wyodrębnić i zidentyfikować sprawców przestępczego działania. R. Abbott, A. Sarch: *Punishing Artificial Intelligence: Legal Fiction or Science Fiction*. "UC Davis Law Review" 2019, vol. 53, s. 375.

³¹ D. Duraes, P. Miguel Freitas, P. Novais: *The Relevance of Deepfakes in the Administration of Criminal Justice*. W: *Multidisciplinary Perspectives on Artificial Intelligence and the Law*. Eds. H.S. Antunes i in. Springer, Cham 2024, s. 366.

w rękach programisty lub użytkownika, któremu z uwagi na brak zdolności umysłowych człowieka odpowiedzialności przypisać nie można. Z kolei w przypadku użytkownika i programisty występuje wewnętrzny element odzwierciedlający się w umyśle pod postacią zamiaru (programista zamierza dokonać podpalenia, a użytkownik chce napaści). Samo zaś wykorzystanie sztucznej inteligencji nie wpływa na odpowiedzialność SI, która według tego modelu jej nie ponosi³². Przekładając to na grunt polski, przykładem takiego zastosowania są np.: odpowiedzialność karna za plagiat i przywłaszczenie autorstwa (art. 303 Prawa własności przemysłowej); odpowiedzialność za uzyskanie przez sprawcę bez uprawnienia dostępu do informacji (art. 267 k.k.); odpowiedzialność za bezprawne wytwarzanie, pozyskiwanie, zbywanie lub udostępnianie programów komputerowych (art. 269b k.k.); odpowiedzialność za oszustwo komputerowe (art. 287 k.k.); odpowiedzialność za uporczywe nękanie (art. 190a § 2 k.k.).

Model odpowiedzialności za popełnione przestępstwo z naturalnym prawdopodobieństwem to model, który uwzględnia wysokie zaangażowanie zarówno programisty systemu SI, jak i jego użytkownika w codzienną pracę sztucznej inteligencji i odnosi się do prawdopodobieństwa wystąpienia skutku w postaci przestępczego działania (pomimo braku wewnętrznego elementu w umysłach programisty czy użytkownika bądź planowania przez nich przestępczego działania). Taką sytuację odzwierciedlają użycia systemów sztucznej inteligencji w pełni autonomicznych. Zarówno użytkownik, jak i osoba je programująca nie mają kontroli nad aktywnościami podejmowanymi przez system AI – nie wiedzą o przestępstwie, nie planują go ani nie mają zamiaru jego dokonania. Zgodnie z tym modelem odpowiedzialność wynikać będzie z możliwości przewidzenia wystąpienia skutku, który jako naturalną konsekwencję podjęcia danych aktywności należało przewidzieć. W takim przypadku programistę i użytkownika uznaje się za gwarant niewystąpienia skutku z uwagi na ich prawny obowiązek zapobiegania kryminalnym następstwom czynów tych podmiotów, które wykazywały standardową aktywność i konsekwencje ich działań można było przewidzieć. Osoby te odpowiedzą jak za zaniedbanie. Na płaszczyźnie polskiego kodeksu karnego w takiej sytuacji aktualizuje się zastosowanie art. 2 k.k.

Ostatni model odpowiedzialności karnej, model bezpośredniej odpowiedzialności systemu sztucznej inteligencji, abstrahuje od zachowania programisty i użytkownika. Odpowiedzialność karnopravna składa się z aspektu wewnętrznego i zewnętrznego. O ile nie stanowi problemu wyodrębnienie zewnętrznego elementu odpowiedzialności systemu sztucznej

³² G. Hallevy: *The Basic Models of Criminal Liability of AI Systems and Outer Circles*. Springer, Berlin 2019, s. 3.

inteligencji (np. uderzenie przez bioniczne ramię, uszkodzenie tkanek przez robota operacyjnego, potrącenie pieszego przez pojazd autonomiczny, powstrzymanie się od określonego działania w przypadku przestępstwa z zaniechania), sytuacja zmienia się w złożoną w razie konieczności udowodnienia istnienia wewnętrznego elementu w postaci zamiaru. Nie ma problemu z przypisaniem systemowi sztucznej inteligencji wiedzy czy możliwości odbierania przez rozmaite receptory dźwięków, dotyku i obrazów. System sztucznej inteligencji wraz ze swymi bogatymi zbiorami danych i zdolnością do samouczenia się już teraz przewyższa zdolności analityczne człowieka. Procesy analizy maszyny i człowieka nie różnią się między sobą znacznie. System sztucznej inteligencji ma do wykonania rozliczne zadania określone jako konkretne cele. Przyjmując teorię instrumentalną, uznającą SI za narzędzie w rękach człowieka, nie można twierdzić, że system sztucznej inteligencji posługuje się swoim wewnętrznym systemem motywacyjnym – działa bowiem w obrębie komend wydawanych przez programistę lub użytkownika.

W piśmiennictwie twierdzi się, że nic nie stoi na przeszkodzie, by dopuścić i uwzględnić możliwość wypełnienia przesłanki wewnętrznej przez system sztucznej inteligencji w sytuacji zamiaru ewentualnego niedbalstwa, działania na zasadzie ryzyka³³. Jednocześnie może dojść do zwolnienia z odpowiedzialności *per analogiam* do odpowiedzialności nieletnich czy osób chorych psychicznie z uwagi na brak możliwości przypisania winy. Wydaje się jednak, że samoucząca się sztuczna inteligencja bardzo dobrze rozróżnia czyny dozwolone od niedozwolonych. Przyjmując model rozsądnego programisty i rozsądnego użytkownika, trzeba stwierdzić, że przy wykonywaniu wszelkich obowiązków, jak np. aktualizacja oprogramowania w przypadku programisty czy poprawne korzystanie z niego w oparciu o regulamin użytkownika systemu SI, należałoby oczekiwać, że do czynu zabronionego nie dojdzie. W takim przypadku programista i użytkownik odpowiadają za konkretne przestępstwo popełnione przez zaniechanie³⁴.

Według G. Hallevy'ego nie ma podstaw do ograniczenia odpowiedzialności samego systemu sztucznej inteligencji w momencie, gdy takiej odpowiedzialności nie ponoszą programista i użytkownik, bądź tylko ją współdzielą. Systemy sztucznej inteligencji posiadają zarówno zdolności faktyczne, jak i zdolności psychiczne pod postacią możliwości rozróżnienia zła od dobra³⁵. Do systemu AI nie będzie miała zastosowania sytuacja spożywania alkoholu, ale proponuje się, by równolegle

³³ Ibidem, s. 7, 11.

³⁴ Ibidem.

³⁵ Ibidem.

na zasadzie analogii stosować taki przepis np. do sytuacji zainfekowania danego oprogramowania wirusem. Według autora do sztucznej inteligencji można wprost stosować przepisy odnoszące się do kontratypów w formie obrony koniecznej i stanu wyższej konieczności³⁶.

Ciekawą propozycją jest również piramida sankcji karnych w kontekście AI przedstawiona przez Fincana Müslüma. U szczytu piramidy znajduje się całkowite i stałe wyłączenie systemu AI jako najdotkliwsza kara, analogiczna do kary śmierci. Nieco niżej jest tymczasowe wyłączenie systemu AI, do czasu znalezienia rozwiązania i wyeliminowania awarii. Dalej mieści się naprawa szkód publicznych. Ostatnia sankcja to proces naprawy szkód poprzez przeprogramowanie systemu AI³⁷.

Jednocześnie trzeba podkreślić, że nieprzewidywalność sztucznej inteligencji ogranicza standardowe zachowania, za które mógłby odpowiadać ktoś inny niż sama AI. Należy też spodziewać się pewnych przestępstw, w przypadku których nie będzie możliwe wskazanie ich sprawcy (dotyczy to zwłaszcza systemów sztucznej inteligencji o dużym stopniu autonomiczności).

Opisane modele nie muszą być stosowane alternatywnie. Mogą się łączyć albo występować niezależnie i pojedynczo. Może być również tak, że sprawcą bezpośrednim będzie człowiek, który odpowie według modelu bezpośredniej odpowiedzialności, a równolegle możliwe będzie przypisanie pomocnictwa systemowi AI, który przysposobił środki i zaplanował sprzeczne z prawem działanie człowieka. Te trzy modele współtworzą siatkę odpowiedzialności w odniesieniu zarówno do ludzi, jak i do systemu sztucznej inteligencji, nie pozwalając tym samym na pominięcie któregośkolwiek ze sprawców odpowiedzialności karnej.

5. Kilka uwag na temat przyjęcia odpowiedzialności karnej za system AI na gruncie polskiego prawa karnego – rozważania *de lege lata*

Na ten moment w polskim systemie prawa karnego negatywnie należy ocenić możliwość zastosowania modelu bezpośredniej odpowiedzialności systemu sztucznej inteligencji. Przyjęcie osobowości prawnej sztucznej inteligencji dla prawa karnego oznacza zmiany podstaw prawa i próbę rozszerzenia stosowania art. 2 k.k. na inne podmioty, takie jak sztuczna inteligencja. „Osobowość prawna jest niezbędna do oskarżenia i skazania AI za przestępstwo, a nadanie AI osobowości prawnej stworzyłoby zupeł-

³⁶ Ibidem.

³⁷ F. Müslüm: *Artificial Intelligence and Legal Issues: A Review of AI-based Legal Impasses in Terms of Criminal Law*. Duncker & Humblot GmbH, Berlin 2022.

nie nowy tryb odpowiedzialności karnej, podobnie jak odpowiedzialność karna korporacji stanowi nowy taki tryb wykraczający poza indywidualną odpowiedzialność karną³⁸. Kluczowym warunkiem wstępnym osobowości prawnej jest pojawienie się samoświadomości, która pozwala na to, by traktować byt jako odpowiedzialnego agenta zdolnego do refleksji nad swoimi działaniami jako własnymi, co stanowi warunek wstępny celowego działania³⁹. Zdecydowana większość doktryny prawa opowiada się przeciwko przyznaniu osobowości prawnej systemowi sztucznej inteligencji, co wyklucza możliwość przyjęcia osobowości prawnej maszyn. Odpowiedzialność prawnokarna za roboty i systemy AI nie jest obecnie możliwa⁴⁰ ze względu na brak wymaganych właściwości podmiotu, który może być pociągnięty do odpowiedzialności⁴¹.

Przypisanie odpowiedzialności karnej systemom sztucznej inteligencji na gruncie polskiego kodeksu karnego nie znajduje zastosowania do sztucznej inteligencji. Nie jest możliwe uznanie systemu SI za sprawcę czynu zabronionego – może być nim tylko człowiek. Należałoby dokonać gruntownej zmiany przepisów karnych odnoszących się do definicji takich pojęć jak: wina, wola, czyn, świadomość, a może nawet skonstruować nowe *lex informatica criminalis*.

Na gruncie polskiej doktryny prawa karnego wina nie jest pojęciem ekwiwalentnym. Psychologiczna teoria winy upatruje jej istotę w stosunku psychicznym sprawcy czynu, a zatem wina jest równoznaczna z przeżyciem psychicznym. Koncepcja ta wywodzi się z kantyzmu i oznacza, że „wola jest rodzajem przyczynowości istot żyjących, o ile są rozumne, a wolność byłaby tą własnością tej przyczynowości, dzięki której może ona działać niezależnie od obcych skłaniających ją przyczyn: tak jak przyrodzona konieczność jest właściwością wszystkich istot bezrozumnych, że do

³⁸ R. Abbott, A. Sarch: *Punishing Artificial Intelligence...*, s. 375. Za odrzuceniem możliwości przypisania przymiotu osobowości prawnej opowiadają się: S. Jankowska: *Podmiotowość prawna sztucznej inteligencji?*. W: *O czym mówią prawnicy, mówiąc o podmiotowości*. Red. A. Bielska-Brodziak. Wydawnictwo Uniwersytetu Śląskiego, Katowice 2021, s. 191; S. Chopra, L.F. White: *A Legal Theory for Autonomous Artificial Agents*. University of Michigan Press, Ann Arbor 2011, s. 185–186; B. Brożek, M. Jakubiec: *On the Legal Responsibility of Autonomous Machines*. „Artif Intell Law” 2017, vol. 25, s. 303; M. Świerczyński, Z. Więckowski: *Sztuczna inteligencja w prawie międzynarodowym. Rekomendacje wybranych rozwiązań*. Difin, Warszawa 2021, s. 74; S.M.C. Avila Negri: *Robot as Legal Person: Electronic Personhood in Robotics and Artificial Intelligence*. „Frontiers in Robotics and AI” 2021, vol. 8, s. 8; M.U. Scherer: *Of Wild Beasts and Digital Analogues: The Legal Status of Autonomous Systems*. „Nevada Law Journal” 2019, no. 1, s. 262.

³⁹ U. Pagallo: *The Laws of Robots: Crimes, Contracts, and Torts*. Springer, Berlin 2013.

⁴⁰ R. Abbott, A. Sarch: *Punishing Artificial Intelligence...*, s. 332.

⁴¹ K. Mamak: *Robotics, AI and Criminal Law: Crimes Against Robots*. Routledge, New York 2024.

działania skłania je wpływ obcych przyczyn”⁴². W stosunku do systemów sztucznej inteligencji koncepcję tę należy odrzucić.

Normatywna teoria winy zasada się na zarzucalnym stosunku psychicznym sprawcy do czynu. Do stwierdzenia winy konieczne jest zatem postawienie sprawcy zarzutu⁴³. Aktualizacja zarzucalności ma miejsce wtedy, gdy możemy ocenić zachowanie sprawcy jako poczytalne oraz gdy zachodzą normalne warunki, w których podejmuje on swe działania⁴⁴. Nie do końca jednak wiadomo, jak mierzyć „poczytalność” systemu sztucznej inteligencji. Prowadzone niejako na marginesie, ale związane z podobnymi wątpliwościami co mierzenie „poczytalności” systemu AI są rozważania oscylujące wokół pojęcia wieku systemu AI. Nie sposób ustalić przełożenie pojęcia „nieletniość” dla sztucznej inteligencji. Pojawia się bowiem pytanie: jak należy liczyć wiek AI? Od pierwszego uruchomienia? Od wprowadzenia do obrotu?

Z kolei kompleksowa teoria winy za podstawę uznaje zarzut stawiany sprawcy czynu, który jest bezprawny, karalny i karygodny⁴⁵. Andrzej Wąsek, będący zwolennikiem kompleksowej teorii winy, na gruncie Kodeksu karnego z 1997 r. definiuje winę jako zarzucalną umyślność lub nieumyślność sprawcy czynu zabronionego⁴⁶. Również ta koncepcja nie znajduje zastosowania do systemu sztucznej inteligencji, jako że jej elementem jest stosunek psychiczny, którego – jak wyżej wskazano – sztuczna inteligencja nie posiada.

Należy także rozważyć zastosowanie do systemu sztucznej inteligencji teorii relacji reprezentowanej przez Agnieszkę Barczak-Oplustil, według której „Relacja bycia winnym zostaje [...] spełniona jeżeli sprawca w chwili czynu charakteryzuje się takim zespołem cech czy właściwości (posiada taką subiektywną kondycję), która pozwala mu na zaniechanie danego postępowania”⁴⁷. Również tę koncepcję należy ocenić negatywnie, jako

⁴² Ł. Kurek: *Problem wolnej woli z perspektywy nauk empirycznych*. „Logos i Ethos” 2011, z. 30, s. 119–120.

⁴³ K. Buchała, w: *Wina w prawie karnym*. Red. W. Wolter. Państwowe Wydawnictwo Naukowe, Kraków 1954, s. 48.

⁴⁴ A. Zoll: *O wartościowaniu czynu w prawie karnym (Uwagi na marginesie artykułu prof. I. Andrejewa o pojęciu winy)*. „Państwo i Prawo” 1983, z. 4, s. 110.

⁴⁵ Z. Jędrzejewski: *Bezprawność jako element przestępności czynu. Studium na temat struktury przestępstwa*. Wolters Kluwer, Warszawa 2009; G. Rejman: *Teorie i formy winy w prawie karnym*. Wydawnictwo Prawnicze, Warszawa 1980, s. 30.

⁴⁶ A. Wąsek: *Ewolucja pojęcia winy w prawie karnym*. „Problemy Prawa Karnego” 1990, z. 4, s. 6. „Winą jest ujemnie oceniona (zarzucalna) umyślność lub nieumyślność”. K. Buchała i in.: *Komentarz do kodeksu karnego. Część ogólna*. Wydawnictwo Prawnicze, Warszawa 1994, s. 39.

⁴⁷ A. Barczak-Oplustil: *Sporne zagadnienia istoty winy w prawie karnym. Zarys problemu*. „Czasopismo Prawa Karnego i Nauk Penalnych” 2005, z. 2 s. 86.

że przypisanie odpowiedzialności karnej następuje w sytuacji, w której sprawca (tu: AI) podejmuje decyzję, mając świadomość oceny następstw swego działania lub zaniechania (legitymował się wewnętrzną i subiektywną możliwością dochowania określonego nakazu lub zakazu, wynikających z danej normy sankcjonowanej).

Podsumowując ten fragment analizy – żadna z zaprezentowanych teorii winy nie znajduje zastosowania do sztucznej inteligencji.

W tym miejscu należy zastanowić się również, czy działanie systemu sztucznej inteligencji mieści się w definicji czynu. Z tego względu warto przywołać kilka najbardziej popularnych koncepcji w tym przedmiocie. Pierwsza za czyn uznaje każde zewnętrzne działanie człowieka, które jest wyrazem jego wolnej woli (wyboru) co do postępowania, które posiada wartość społeczną (kontekst). Zwolennikami tego poglądu są: Leszek Lernell, Andrzej Zębik i Andrzej Marek. Trzeba jednak odrzucić tę koncepcję w stosunku do robota sztucznej inteligencji, gdyż nie jest on człowiekiem⁴⁸. Podobną definicję zaproponował Marian Cieślak, który twierdzi, że czyn to wycinek ludzkiego zachowania o społecznej wartości, mający znaczenie dla określonego dobra prawnego⁴⁹. I gdyby sugerować się tylko drugą częścią tej definicji, to można by uznać, że zachowanie systemu sztucznej inteligencji może mieć wartość społeczną dla określonych dóbr prawnych. Nietrudno bowiem wyobrazić sobie roboty społeczne, roboty-asystentów, które pełnią ważne funkcje społeczne. Wciąż jednak nie jest spełniona przesłanka bycia człowiekiem. Inny pogląd mówi o tym, że czyn jest działaniem (sekwencją ruchu) lub zaniechaniem i stanowi „techniczną” stronę prawnie relewantnego zachowania⁵⁰. Z kolei definicję czynu połączonego z ukierunkowaniem go na cel przedstawiają w nauce: Witold Świda, Igor Andrejew, Tadeusz Hanausek i Kryspin Mioduski⁵¹. Żadna

⁴⁸ L. Lernell: *Wykład prawa karnego. Część ogólna*. T. 1. Wydawnictwo Uniwersytetu Warszawskiego, Warszawa 1969, s. 77, 108; A. Zębik, w: *Prawo karne w zarysie. Nauka o ustawie karnej i przestępstwie*. Red. J. Waszczyński. Wydawnictwo Uniwersytetu Łódzkiego, Łódź 1992, s. 132; A. Marek: *Prawo karne*. C.H.Beck, Warszawa 2004, s. 114.

⁴⁹ M. Cieślak: *Pojęcie czynu na tle przesłanek odpowiedzialności karnej w obowiązującym prawie polskim*. „Zeszyty Naukowe Uniwersytetu Gdańskiego” 1987, nr 15, s. 39.

⁵⁰ A. Marek: *Prawo karne...*, s. 97. Dodatkowo warto wspomnieć o koncepcji działania jako kompleksu ruchów i ruchów zahamowanych oraz koncepcji odruchów warunkowych lub braku oczekiwanej aktywności Kazimierza Buchały. K. Buchała: *Prawo karne materialne*. Państwowe Wydawnictwo Naukowe, Warszawa 1980, s. 186; K. Buchała, w: K. Buchała, S. Waltoś: *Zasady prawa i procesu karnego*. PWN, Warszawa 1975, s. 44.

⁵¹ W. Świda: *Prawo karne*. PWN, Warszawa 1978, s. 118; I. Andrejew: *Ustawowe znamiona czynu. Typizacja i kwalifikacja przestępstw*. Wydawnictwo Prawnicze, Warszawa 1978, s. 139; T. Hanausek: *Wyłączenie czynu jako podstawowego elementu przestępstwa*. „Wojskowy Przegląd Prawniczy” 1967, nr 4, s. 435; K. Mioduski, w: J. Bafia, K. Mioduski, M. Siewierski: *Kodeks karny. Komentarz*. Wydawnictwo Prawnicze, Warszawa 1977, s. 12.

z istniejących koncepcji czynu w polskiej doktrynie prawa karnego nie przystaje do systemów sztucznej inteligencji. Należy przychylić się zatem do twierdzenia Lecha Gardockiego, zgodnie z którym „przestępstwem może być tylko czyn, [co – O.K.] przesądza też kwestię ewentualnej odpowiedzialności za przestępstwo podmiotów niebędących osobami fizycznymi. Podmioty takie nie mogą, w świetle definicji przestępstwa, ani przestępstwa popełnić, ani ponieść za nie odpowiedzialności karnej”⁵².

Jak dowiedziono powyżej, główną przeszkodą przypisywalności odpowiedzialności jest wina. Dodatkowo, decyzje podejmowane w ramach procesu „zachowania się” sztucznej inteligencji nie mieszczą się w kategorii czynu⁵³.

6. Alternatywne podejścia Davida J. Gunkela, Svena Nyholma, Ryana Abbotta i Kamila Mamaka

W tej części artykułu chciałabym przedstawić kilka alternatywnych perspektyw odnoszących się do odpowiedzialności karnej za sztuczną inteligencję. David J. Gunkel wprowadził trzy możliwości odpowiedzi na lukę w odpowiedzialności: ponoszenie odpowiedzialności przez ludzi niezależnie od poziomu zaawansowania sztucznej inteligencji (*de facto* jest to pozostawienie stanu obecnego); quasi-odpowiedzialność maszyn; odpowiedzialność rozproszona, polegająca na traktowaniu sztucznej inteligencji jak człowieka, ale z elementami systemu autonomicznego⁵⁴. Ostatnie podejście jest szczególnie propagowane przez Svena Nyholma, który słusznie zauważa, że w kontekście systemów sztucznej inteligencji powinniśmy prawie zawsze myśleć o sprawczości jako występującej w ramach współ-

⁵² L. Gardocki: *Pojęcie przestępstwa i podziały przestępstw w polskim prawie karnym*. „Annales Universitatis Mariae Curie-Skłodowska Lublin-Polonia” 2013, vol. LX, 2, Sectio G, s. 31.

⁵³ Na czyn przestępczy składają się dwa elementy: działanie, czyli popełnienie czynu zabronionego (*actus reus*), oraz zdolność rozpoznawania czynu zabronionego (*mens rea*). Problematyczna jest strona podmiotowa zarówno w prawie anglosaskim (gdzie występuje doktryna *mens rea*), jak i w polskim porządku, gdzie nie da się przypisać winy sztucznej inteligencji. „Alternatywnie można by opracować nową fikcję prawną AI *mens rea*, niejako analogiczną do ludzkiego *mens rea*, ale to również nie jest obecnie wykonalne rozwiązanie. To podejście mogłoby wymagać zeznań biegłych, aby umożliwić sądom szczegółowe rozważenie, w jaki sposób odpowiednia AI działała, aby ocenić, czy była w stanie rozważyć prawnie istotne wartości i interesy, ale nie ważyła ich wystarczająco, oraz czy program ma odpowiednie dyspozycje behawioralne związane z intencją lub wiedzą podobną do *mens rea*”. R. Abbott, A. Sarch: *Punishing Artificial Intelligence...*, s. 375.

⁵⁴ D.J. Gunkel: *Mind the Gap: Responsible Robotics and the Problem of Responsibility*. „Ethics and Information Technology” 2020, vol. 6 (4), s. 284.

pracy człowieka z robotem⁵⁵. Mówi on o swego rodzaju kategoriach zespołów człowiek-technologia i człowiek-maszyna, a nie o kategoriach maszyn lub technologii działających samodzielnie⁵⁶. W związku z tym istoty ludzkie mogą nadal być odpowiedzialne zarówno za dobre, jak i za złe wyniki pracy systemów sztucznej inteligencji⁵⁷.

O nieredukowalnym zachowaniu przestępczym AI i luce w odpowiedzialności pisze również Ryan Abbott. Jego zdaniem należy wprowadzić minimalne rozszerzenie prawa karnego oraz zadbać o poprawne stosowanie w aspekcie używania nowych technologii zasady *crimen sine lege certa* (zachowanie penalizowane powinno być ściśle określone przez prawo⁵⁸, a prawo karne powinno być zrozumiałe dla adresatów⁵⁹). Samo rozszerzenie prawa karnego miałoby się odbyć poprzez zdefiniowanie nowych przestępstw osób fizycznych, wprowadzenie ustawy o nadużyciach AI oraz kryminalizację braku odpowiedniego wdrażania, testowania, projektowania, monitorowania i nadzorowania systemu AI⁶⁰ (np. projektowanie lub obsługa AI zdolnej do wyrządzenia szkody bez wyznaczenia osoby odpowiedzialnej mogłyby być przestępstwem *per analogiam* do prowadzenia pojazdu bez prawa jazdy). Takie rozwiązanie mogłoby być jednak administracyjnie problematyczne z uwagi na konieczność prowadzenia rejestrów osób odpowiedzialnych. W związku z tym R. Abbott proponuje przyjęcie domyślnej reguły, że osobą odpowiedzialną w przypadku produktu komercyjnego jest producent lub dostawca AI, a w przypadku produktu niekomercyjnego – właściciel AI, twórca bądź użytkownik⁶¹. Ta propozycja, szczególnie w świetle przyjęcia AIA, znajduje zastosowanie do odpowiedzialności cywilnoprawnej za działania sztucznej inteligencji.

Ryan Abbott postuluje także wprowadzenie umiarkowanych zmian w zakresie odpowiedzialności cywilnej, przede wszystkim rozciągnięcie odpowiedzialności deliktowej (oraz wprowadzenie systemu ubezpieczeń od AI), którą można by stosować nawet w połączeniu z rozszerzeniem odpowiedzialności karnej, co stanowi kompletne rozwiązanie, jeśli weźmiemy

⁵⁵ S. Nyholm: *Humans and Robots, Ethics, Agency, and Anthropomorphism*. Rowman & Littlefield Publishers, London 2020, s. 72.

⁵⁶ M. Czochra, D. Bar: *Śmierć pacjenta...*, s. 77.

⁵⁷ S. Nyholm: *This is Technology Ethics*. Wiley, Hoboken 2023, s. 179.

⁵⁸ R. Lahti: *The Rule of Law and Finnish Criminal Law Reform*. „Acta Juridica Hungarica. Hungarian Journal of Legal Studies” 1996, vol. 37, s. 251.

⁵⁹ L. Loucaides: *Essays on the Developing Law of Human Rights*. Brill-Nijhoff, Norwell 1995, s. 32.

⁶⁰ R. Abbott, A. Sarch: *Punishing Artificial Intelligence...*, s. 379.

⁶¹ Podobnie wszystkie systemy sztucznej inteligencji mają ludzkich programistów, a w przypadku, gdy oprogramowania tworzy autonomicznie inny system sztucznej inteligencji, należy pociągnąć do odpowiedzialności pierwotnego właściciela sztucznej inteligencji.

pod uwagę wspomniane wcześniej prace Parlamentu Europejskiego zmierzające ku określeniu ram odpowiedzialności cywilnej AI⁶².

Na koniec przeglądu zagranicznej doktryny prawa warto również wspomnieć o Gabrieli Hallevym, który przewiduje możliwość przypisania odpowiedzialności prawnokarnej systemowi AI na zasadzie analogii do odpowiedzialności podmiotów zbiorowych⁶³. Takie rozwiązanie wymagałoby redefinicji podstawowych pojęć: winy, woli, świadomości oraz sposobu karania⁶⁴. W tym miejscu proponuje się stworzenie odrębnej gałęzi prawa, która regulowałaby zasady odpowiedzialności za działania (zaniechania) systemów AI oraz stosunki prawne z jej udziałem⁶⁵.

W polskiej nauce problem zagrożeń związanych z ogólną sztuczną inteligencją (AGI)⁶⁶ zauważa Kamil Mamak, który jako jedno z narzędzi służących regulacji następstw i zagrożeń stwarzanych przez AGI proponuje zastosowanie przepisów karnych oraz ustanowienie przestępstw związanych z AGI. Jeszcze bez określenia wymiaru kary K. Mamak proponuje następujące brzmienie projektu dotyczącego penalizowania zachowań AGI: „§ 1. Każdy, kto wprowadza niekontrolowaną sztuczną inteligencję ogólną, podlega karze. § 2. Jeżeli sprawca czynu, o którym mowa w § 1, działa nieumyślnie, podlega karze. § 3. Kto przygotowuje się do popełnienia przestępstwa, o którym mowa w § 1, podlega karze”⁶⁷. Przymiotnik „niekontrolowany” odnosi się do braku wpływu na kierunek działania AGI lub braku możliwości jej wyłączenia. Z kolei „wprowadzenie” to według autora tej propozycji działanie doprowadzające AGI do niekontrolowanego aktu (może nim być zarówno działanie programistyczne, jak i uruchomienie narzędzia przez dyrektora firmy). Według K. Mamaka samo tworzenie generalnej sztucznej inteligencji nie stanowi przestępstwa, ale jedynie jej

⁶² R. Abbott, A. Sarch: *Punishing Artificial Intelligence...*, s. 381.

⁶³ G. Hallevy: *When Robots Kill: Artificial Intelligence under Criminal Law*. Northeastern University Press, Boston 2013, s. 177.

⁶⁴ R. Rejmank: *Autonomiczność...*, s. 109.

⁶⁵ R. Rejmank: *Odpowiedzialność karna za skutki spowodowane przez autonomiczne roboty bojowe – zarys problemu*. W: *Wykorzystanie dronów i robotów w systemach bezpieczeństwa. Wybrane aspekty*. Red. R. Kamprowski, M. Skarżyński. Wydawnictwo Naukowe WNPiD UAM, Poznań 2019, s. 44.

⁶⁶ AGI zdefiniowano jako formę sztucznej inteligencji, która „dorównuje lub przewyższa ludzką inteligencję w wielu różnych zadaniach poznawczych”. AGI nie istnieje – zgodnie ze stanem dzisiejszej nauki, ale wąska SI może rozwiązać wiele problemów bez konieczności bycia inteligentnym. Idea AGI bazuje na hipotezie, że można poczynić postęp w kierunku zwiększonej i szerszej inteligencji sztucznych agentów. Biorąc pod uwagę tempo, w jakim sztuczna inteligencja została opracowana i przyjęta w skali globalnej, możliwości powstania AGI w przyszłości nie można wykluczyć. T. Mahler: *Regulating Artificial General Intelligence (AGI)*. W: *Law and Artificial Intelligence...*, s. 522.

⁶⁷ K. Mamak: *AGI Crimes? The Role of Criminal Law in Mitigating Existential Risks Posed by Artificial General Intelligence*. „AI & Society” 2025, vol. 40, s. 2697.

niekontrolowana wersja, wobec czego autor postuluje, by twórcy upewnili się, że zanim dany system AI wystartuje, człowiek zawsze będzie mieć możliwość jego kontrolowania⁶⁸. Koncepcje trzech przestępstw, w tym dwóch zamierzonych i jednego niezamierzonego, należy ocenić pozytywnie. Poprzez działanie odstrasżające prawa karnego nowe prawo może stanowić instrument karania także i wobec tych, którzy zamierzają pracować nad złośliwą AGI⁶⁹.

7. Podsumowanie

Rozwiązaniem problemu odpowiedzialności za działanie i zaniechanie sztucznej inteligencji w przyszłości może stać się przyznanie AI osobowości prawnej, co otworzy możliwość pociągnięcia jej do odpowiedzialności karnej. Wydaje się, że same gwarancje instrumentów znanych cywilistycie czy utrzymanie znaczącej kontroli i nadzoru człowieka (choćby ścisła odpowiedzialność właściciela zwierzęcia [art. 431 k.c.] lub odpowiedzialność za nadzór ze strony człowieka nadzoru [art. 430 k.c.]) mogą okazać się dostateczne, dopóki nie zgłębimy potencjału wciąż rozwijającej się sztucznej inteligencji. Zestawienie w artykule porządku anglosaskiego i kontynentalnego prawa karnego wskazuje, że możliwe jest, iż prawo karne nie jest jeszcze gotowe na takie rozwiązania, ani też nie będzie stanowić remedium na problemy związane z odpowiedzialnością za sztuczną inteligencję. Do tego dochodzi taka regulacja nadzoru, która w sposób klarowny wskaże, kto podlega obowiązkowi nadzoru, tak aby zminimalizować mogące potencjalnie wystąpić błędne wyobrażenia na temat odpowiedzialności danego podmiotu. Cywilnoprawny nadzór człowieka można by rozciągnąć na prawo karne. Dopóki jednak nie istnieje superinteligencja, dopóty musi istnieć możliwość unieważnienia decyzji AI lub jej kontroli przez człowieka. Tworzenie i produkowanie takich rozwiązań, które wykluczają człowieka z pętli decyzyjnej, powinno być zabronione. Prawo karne polskie musiałoby także przeddefiniować naturę działania i współodpowiedzialności, dopuszczając do pociągnięcia do odpowiedzialności nieczłowieka. Na ten moment wydaje się to jednak zbyt rewolucyjne. Na obecnym etapie rozwoju nowych technologii i prawa należy przychylić się do poglądu R. Abbotta, że wystarczą niewielkie zmiany w istniejącym porządku prawa karnego wraz z rozszerzonym prawem do odpowiedzialności cywilnej,

⁶⁸ Ibidem.

⁶⁹ Ibidem.

które według niego stanowią lepsze rozwiązanie problemu przestępstw związanych ze sztuczną inteligencją⁷⁰.

W doktrynie coraz częściej zauważa się przenikające się wpływy prawa cywilnego i prawa karnego: „Jeżeli zatem zarówno w prawie karnym, jak i w prawie deliktów wyróżnia się pojęcia bezprawności, winy oraz umyślności/nieumyślności, to nie ma przeszkód do przyjęcia, że na płaszczyźnie obu, wina to ta sama kategoria pojęciowa, a proces jej przypisania jest tożsamy zarówno na potrzeby odpowiedzialności karnej, jak i odszkodowawczej z tytułu czynu niedozwolonego [...] uzasadnione jest mówienie o jednej, tożsamej winie w prawie karnym i prawie deliktów w zakresie dotyczącym odpowiedzialności odszkodowawczej za własny czyn niedozwolony”⁷¹; „[...] prawo cywilne, konstruując własne pojęcie winy, nie może ujmować go w sposób jakościowo różny od przyjmowanego w obrębie prawa karnego, a pewne różnice między tymi dziedzinami nie mogą prowadzić do zatarcia podstawowej wspólnoty pojęcia winy na całym obszarze prawa”⁷².

W nauce postuluje się rozszerzenie prawa karnego o nowe przestępstwa z tytułu zaniedbania, skupione wokół niewłaściwego projektowania, testowania i obsługi systemu sztucznej inteligencji (w tym miejscu należałoby zdefiniować obowiązki opieki i nadzoru dotyczące sztucznej inteligencji), oraz rozszerzenie odpowiedzialności cywilnej, która mogłaby uzupełniać te ramy⁷³. Przyjmuje się, że modelową odpowiedzialnością mogłaby być ta w reżimie odpowiedzialności za produkt⁷⁴. Dodatkowo R. Abbott proponuje, by w odniesieniu do AI posługiwać się różnego rodzaju fikcjami w celu uelastyczenia prawa karnego⁷⁵. Przyjęcie współsprawstwa na ten moment jest jednak zbyt rewolucyjne.

Jakie to może mieć przełożenie na polskie prawo karne? Być może warto zastanowić się nad dodaniem nowego rozdziału do ustawy – pod postacią nowego katalogu przestępstw związanych z niedbalstwem dotyczącym niewłaściwego projektowania, wdrażania, testowania, szkolenia, monitorowania, obsługi i testowania rozwiązań związanych ze sztuczną inteligencją (R. Abbott mówi o przestępstwie opracowywania systemów, które w sposób przewidywalny mogą stanowić ryzyko powstania poważ-

⁷⁰ Ibidem.

⁷¹ M. Zelek: *Wina w prawie karnym i prawie deliktów – przyczynek do dyskusji na temat tożsamości pojęcia winy w prawie polskim*. „Acta Iuris Stetinensis” 2019, nr 2 (26), s. 124.

⁷² Ibidem, s. 110. Podobnie: M. Sośniak: *Bezprawność zachowania jako przesłanka odpowiedzialności cywilnej za czyny niedozwolone*. Uniwersytet Jagielloński, Kraków 1959, s. 81; J. Dąbrowa: *Wina jako przesłanka odpowiedzialności cywilnej*. Zakład Narodowy im. Ossolińskich, Wrocław 1968, s. 141.

⁷³ R. Abbott, A. Sarch: *Punishing Artificial Intelligence...*, s. 382.

⁷⁴ Ibidem.

⁷⁵ Ibidem.

nej szkody lub konsekwencji niezgodnych z prawem, nawet jeżeli istniało szczególne ryzyko, którego powstania nie dało się przewidzieć)⁷⁶. Jest to jednak ryzykowne posunięcie, gdyż *de facto* każde oprogramowanie może wiązać się z wystąpieniem jakiegoś stopnia ryzyka. Takie rozwiązanie mogłoby powodować zaniechanie innowacji. To również problem związany z nakładaniem sankcji karnych na osoby, które nie dopełniły obowiązku nadzoru nad systemem sztucznej inteligencji⁷⁷. Rozbudowane podstawy odpowiedzialności cywilnej zawarte w DPLD i AILD mogłyby stanowić uzupełnienie ram odpowiedzialności za sztuczną inteligencję.

Na podstawie obowiązującego Kodeksu karnego odpowiedzialność karłą można przypisać wyłącznie człowiekowi. Przedstawione powyżej teorie winy oraz definicje czynu nie są adekwatne do działań i zaniechań systemu sztucznej inteligencji. Aby w ogóle można było móc mówić o odpowiedzialności samej sztucznej inteligencji, należałoby skonstruować odrębne przesłanki odpowiedzialności, uwzględniające specyfikę nowych technologii. Taka sytuacja rodzi rozmaite wątpliwości dotyczące ewentualnego zastosowania przepisów o niepoczytalności, stanu wyższej konieczności i obrony koniecznej, a także tego, czy w ogóle istniałaby możliwość ich stosowania do AI oraz celów kary (wychowawcze, prewencyjne). W przypadku przestępstwa przy użyciu systemu sztucznej inteligencji można skupić się na szukaniu osoby fizycznej w całym łańcuchu odpowiedzialności (np. producenta, programisty czy importera). Pociągnięcie do odpowiedzialności karnej samej sztucznej inteligencji nie jest i najprawdopodobniej jeszcze przez jakiś czas nie będzie możliwe.

Prawo karne jest niewystarczające w przypadkach, gdy sztuczna inteligencja powoduje szkody, w których nie sposób wyodrębnić i zidentyfikować sprawców przestępczego działania⁷⁸. Ostatecznie przyjmuje się, że karanie systemu AI nie jest uzasadnione (przyjęcie takiego rozwiązania wiązałoby się z dużymi kosztami, problemami praktycznymi i daleko posuniętymi zmianami prawa karnego)⁷⁹. W polskim porządku prawnym nie jest możliwe przypisanie odpowiedzialności prawnokarnej sprawcy niebędącemu człowiekiem⁸⁰. Jeżeli przyjmiemy kontrolę człowieka nad rozwojem i wdrażaniem sztucznej inteligencji, będziemy w stanie tolerować nieuniknione błędy i unikać ich w przyszłości⁸¹. I choć nie z każdym

⁷⁶ Ibidem.

⁷⁷ Ibidem.

⁷⁸ Ibidem, s. 381.

⁷⁹ Ibidem.

⁸⁰ R. Rejmaniak: *Autonomiczność...*, s. 108.

⁸¹ D.L. Gogoshin: *A Way Forward for Responsibility in the Age of AI*. Inquiry, London 2024.

twierdzeniem G. Hallevy’ego można zgodzić się, nieuchronne jest to, że prawo karne musi dostosować się do nowej rzeczywistości⁸².

Bibliografia

Literatura

- Abbott R., Sarch A.: *Punishing Artificial Intelligence: Legal Fiction or Science Fiction*. "UC Davis Law Review" 2019, vol. 53, s. 323–383.
- AI in eHealth. *Human Autonomy, Data Governance and Privacy in Healthcare*. Eds. M.C. Compagnucci, M.L. Wilson, M. Fenwick, M. Forgó, T. Bärnighausen. Cambridge University Press, Cambridge 2022.
- Andrejew I.: *Ustawowe znamiona czynu. Typizacja i kwalifikacja przestępstw*. Wydawnictwo Prawnicze, Warszawa 1978.
- Avila Negri S.M.C.: *Robot as Legal Person: Electronic Personhood in Robotics and Artificial Intelligence*. "Frontiers in Robotics and AI" 2021, vol. 8, s. 1–10.
- Bafia J., Mioduski K., Siewierski M.: *Kodeks karny. Komentarz*. Wydawnictwo Prawnicze, Warszawa 1977.
- Baker D.J., Robinson P.H.: *Artificial Intelligence and the Law: Cybercrime and Criminal Liability*. Routledge, New York 2021.
- Barczak-Oplustil A.: *Sporne zagadnienia istoty winy w prawie karnym. Zarys problemu*. „Czasopismo Prawa Karnego i Nauk Penalnych” 2005, z. 2, s. 79–96.
- Beckers A., Teubner G.: *Three Liability Regimes for Artificial Intelligence*. Hart Publishing, Oxford 2021.
- Brożek B., Jakubiec M.: *On the Legal Responsibility of Autonomous Machines*. "Artif Intell Law" 2017, vol. 25, s. 293–304.
- Buchała K.: *Prawo karne materialne*. Państwowe Wydawnictwo Naukowe, Warszawa 1980.
- Buchała K., Cwiakalski Z., Szewczyk M., Zoll A.: *Komentarz do kodeksu karnego. Część ogólna*. Wydawnictwo Prawnicze, Warszawa 1994.
- Buchała K., Waltoś S.: *Zasady prawa i procesu karnego*. PWN, Warszawa 1975.
- The Cambridge Handbook of Responsible Artificial Intelligence. Interdisciplinary Perspectives*. Eds. S. Voeneke, Ph. Kellmeyer, O. Mueller, W. Burgard. Cambridge University Press, Cambridge 2022.
- Chopra S., White L.F.: *A Legal Theory for Autonomous Artificial Agents*. University of Michigan Press, Ann Arbor 2011.
- Cieślak M.: *Pojęcie czynu na tle przesłanek odpowiedzialności karnej w obowiązującym prawie polskim*. „Zeszyty Naukowe Uniwersytetu Gdańskiego” 1987, nr 15, s. 33–44.
- Coeckelbergh M.: *Virtual Moral Agency, Virtual Moral Responsibility: On the Moral Significance of the Appearance, Perception, and Performance of Artificial Agents*.

⁸² G. Hallevy: *Liability for Crimes Involving Artificial Intelligence Systems*. Springer, Cham 2015, s. 229.

- “AI and Society” 2009, vol. 24, s. 181–189. <https://doi.org/10.1007/s00146-009-0208-3>.
- Czochra M., Bar D.: *Śmierć pacjenta wywołana zastosowaniem sztucznej inteligencji w technologiach medycznych – analiza prawnokarna*. „Studia Prawnicze. Rozprawy i Materiały” 2019, nr 2 (25), s. 67–79.
- Dąbrowa J.: *Wina jako przesłanka odpowiedzialności cywilnej*. Zakład Narodowy im. Ossolińskich, Wrocław 1968.
- Dignum V.: *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Springer, Cham 2019.
- Duraes D., Miguel Freitas P., Novais P.: *The Relevance of Deepfakes in the Administration of Criminal Justice*. W: *Multidisciplinary Perspectives on Artificial Intelligence and the Law*. Eds. H.S. Antunes, P.M. Freitas, A.L. Oliveira, C.M. Pereira, E. Vaz de Sequeira, L.B. Xavier. Springer, Cham 2024, s. 351–369.
- Ethics of Socially Disruptive Technologies: An Introduction*. Eds. I. van de Poel, L. Frank, J. Hermann, J. Hopster, D. Lenzi, S. Nyholm, B. Taebi, E. Ziliotti. Cambridge University Press, Cambridge 2023.
- Gardocki L.: *Pojęcie przestępstwa i podziały przestępstw w polskim prawie karnym*. „Annales Universitatis Mariae Curie-Skłodowska Lublin-Polonia” 2013, vol. LX, 2, Sectio G, s. 29–40.
- Gogoshin D.L.: *A Way Forward for Responsibility in the Age of AI*. Inquiry, London 2024.
- Gunkel D.: *How to Survive a Robot Invasion. Rights, Responsibility, and AI*. Routledge, New York 2020.
- Gunkel D.: *Robot Rights*. The MIT Press, Cambridge 2018.
- Gunkel D.J.: *Mind the Gap: Responsible Robotics and the Problem of Responsibility*. “Ethics and Information Technology” 2020, vol. 6 (4), s. 279–290.
- Hallevy G.: *The Basic Models of Criminal Liability of AI Systems and Outer Circles*. Springer, Berlin 2019.
- Hallevy G.: *Liability for Crimes Involving Artificial Intelligence Systems*. Springer, Cham 2015.
- Hallevy G.: *When Robots Kill: Artificial Intelligence under Criminal Law*. Northeastern University Press, Boston 2013.
- Hanausek T.: *Wyłączenie czynu jako podstawowego elementu przestępstwa*. „Wojсковy Przegląd Prawniczy” 1967, nr 4, s. 430–459.
- Hart H.L.A.: *Punishment and Responsibility: Essays in the Philosophy of Law*. 2nd ed. Oxford University Press, Oxford 2008.
- Hawking S.: *Transcendence Looks at the Implication of Artificial Intelligence – But Are We Taking AI Seriously Enough?*. “The Independent”, 01.05.2014. <https://www.independent.co.uk/news/science/stephen-hawking-transcendence-looks-at-the-implications-of-artificial-intelligence-but-are-we-taking-ai-seriously-enough-9313474.html> [dostęp: 09.07.2024].
- Jankowska M.: *Podmiotowość prawna sztucznej inteligencji?*. W: *O czym mówią prawnicy, mówiąc o podmiotowości*. Red. A. Bielska-Brodziak. Wydawnictwo Uniwersytetu Śląskiego, Katowice 2021, s. 171–196.

- Jędrzejewski Z.: *Bezprawność jako element przestępności czynu. Studium na temat struktury przestępstwa*. Wolters Kluwer, Warszawa 2009.
- Kolfschooten H. van: *EU Regulation of Artificial Intelligence, Challenges for Patients' Rights*. "Amsterdam Law School Legal Studies Research Paper, Kluwer Law International" 2021, vol. 59, s. 81–112.
- Kurek Ł.: *Problem wolnej woli z perspektywy nauk empirycznych*. „Logos i Ethos” 2011, z. 30, s. 99–138.
- Lahti R.: *The Rule of Law and Finnish Criminal Law Reform*. „Acta Juridica Hungarica. Hungarian Journal of Legal Studies” 1996, vol. 37, s. 251–258.
- Lernell L.: *Wykład prawa karnego. Część ogólna*. T. 1. Wydawnictwo Uniwersytetu Warszawskiego, Warszawa 1969.
- Loucaides L.: *Essays on the Developing Law of Human Rights*. Brill-Nijhoff, Norwell 1995.
- Mahler T.: *Regulating Artificial General Intelligence (AGI): W: Law and Artificial Intelligence: Regulating AI and Applying AI in Legal Practice*. Eds. B. Custers, E. Fosch-Villaronga. Springer, Berlin 2022, s. 521–540.
- Mamak K.: *AGI Crimes? The Role of Criminal Law in Mitigating Existential Risks Posed by Artificial General Intelligence*. „AI & Society” 2025, vol. 40, s. 2691–2701.
- Mamak K.: *Robotics, AI and Criminal Law: Crimes Against Robots*. Routledge, New York 2024.
- Marek A.: *Prawo karne*. C.H.Beck, Warszawa 2004.
- Matthias A.: *The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata*. „Ethics and Information Technology” 2004, no. 6 (3), s. 175–183.
- Müslüm F.: *Artificial Intelligence and Legal Issues: A Review of AI-based Legal Impasses in Terms of Criminal Law*. Duncker & Humblot GmbH, Berlin 2022.
- Nyholm S.: *Humans and Robots, Ethics, Agency, and Anthropomorphism*. Rowman & Littlefield Publishers, London 2020.
- Nyholm S.: *This is Technology Ethics*. Wiley, Hoboken 2023.
- Pagallo U.: *The Laws of Robots: Crimes, Contracts, and Torts*. Springer, Berlin 2013.
- Picard R.: *Teaching Robots Right from Wrong*. „The Economist”, 05.07.2017. <https://www.economist.com/1843/2017/07/05/teaching-robots-right-from-wrong> [dostęp: 09.07.2024].
- Prawo karne w zarysie. Nauka o ustawie karnej i przestępstwie*. Red. J. Waszczyński. Wydawnictwo Uniwersytetu Łódzkiego, Łódź 1992.
- Prifti K., Stamhuis E., Heine K.: *Digging into the Accountability Gap: Operator's Civil Liability in Healthcare AI-systems*. W: *Law and Artificial Intelligence: Regulating AI and Applying AI in Legal Practice*. Eds. B. Custers, E. Fosch-Villaronga. Springer, Berlin 2022, s. 279–295.
- Quattrocchio S.: *Artificial Intelligence, Computational Modelling and Criminal Proceedings: A Framework for A European Legal Discussion*. Springer, Cham 2020.
- Regulating Artificial Intelligence in Industry*. Ed. D.M. Bielicki. Routledge, New York 2021.

- Rejman G.: *Teorie i formy winy w prawie karnym*. Wydawnictwo Prawnicze, Warszawa 1980.
- Rejmaniak R.: *Autonomiczność systemów sztucznej inteligencji jako wyzwanie dla prawa karnego*. „Roczniki Nauk Prawnych” 2021, t. 31, nr 3, s. 95–113.
- Rejmaniak R.: *Odpowiedzialność karna za skutki spowodowane przez autonomiczne roboty bojowe – zarys problemu*. W: *Wykorzystanie dronów i robotów w systemach bezpieczeństwa. Wybrane aspekty*. Red. R. Kamprowski, M. Skarżyński. Wydawnictwo Naukowe WNPiD UAM, Poznań 2019, s. 95–113.
- Santoni de Sio F., Mecacci G.: *Four Responsibility Gaps with Artificial Intelligence: Why they Matter and How to Address them*. „Philosophy & Technology” 2021, vol. 34, s. 1057–1084. <https://doi.org/10.1007/s13347-021-00450-x>.
- Scherer M.U.: *Of Wild Beasts and Digital Analogues: The Legal Status of Autonomous Systems*. „Nevada Law Journal” 2019, no. 1, s. 259–290.
- Solaiman S.M.: *Legal Personality of Robots, Corporations, Idols and Chimpanzees: A Quest for Legitimacy*. „Artificial Intelligence and Law” 2017, no. 25 (2), s. 155–179. <https://doi.org/10.1007/s10506-016-9192-3>.
- Sośniak M.: *Bezprawność zachowania jako przesłanka odpowiedzialności cywilnej za czyny niedozwolone*. Uniwersytet Jagielloński, Kraków 1959.
- Świda W.: *Prawo karne*. PWN, Warszawa 1978.
- Świerczyński M., Więckowski Z.: *Sztuczna inteligencja w prawie międzynarodowym. Rekomendacje wybranych rozwiązań*. Difin, Warszawa 2021.
- Wąsek A.: *Ewolucja pojęcia winy w prawie karnym*. „Problemy Prawa Karnego” 1990, z. 4, s. 5–15.
- Wina w prawie karnym*. Red. W. Wolter. Państwowe Wydawnictwo Naukowe, Kraków 1954.
- Zelek M.: *Wina w prawie karnym i prawie deliktów – przyczynek do dyskusji na temat tożsamości pojęcia winy w prawie polskim*. „Acta Iuris Stetinensis” 2019, nr 2 (26), s. 109–126.
- Zoll A.: *O wartościowaniu czynu w prawie karnym (Uwagi na marginesie artykułu prof. I. Andrejewa o pojęciu winy)*. „Państwo i Prawo” 1983, z. 4, s. 109–120.

Akty prawne

- Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r. w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji oraz zmiany rozporządzeń (WE) nr 300/2008, (UE) nr 167/2013, (UE) nr 168/2013, (UE) 2018/858, (UE) 2018/1139 i (UE) 2019/2144 oraz dyrektyw 2014/90/UE, (UE) 2016/797 i (UE) 2020/1828 (akt w sprawie sztucznej inteligencji).
- Ustawa z dnia 6 czerwca 1997 r. – Kodeks karny. T.j. Dz.U. 1997, nr 88, poz. 553.
- Ustawa z dnia 30 czerwca 2000 r. – Prawo własności przemysłowej. T.j. Dz.U. 2001, nr 49, poz. 1170.