Ius Novum
2026, vol. 20, nr 1, s. 1–26

WOLNOŚĆ WOLI JAKO PRZESŁANKA ODPOWIEDZIALNOŚCI KARNEJ: CZŁOWIEK I SZTUCZNA INTELIGENCJA W PERSPEKTYWIE CZESKIEGO PRAWA KARNEGO[1]

JAROSLAV FENYK*

DOI: 10.26399/iusnovum.v20.1.2026.01/j.fenyk

Streszczenie
W artykule poddano analizie formę i miejsce woli człowieka w podstawach odpowiedzialności karnej. Wyjaśniono w nim konieczność odpowiedzialności karnej osób fizycznych w sposób niezależny od woli oraz porównano z możliwą odpowiedzialnością karną systemów autonomicznych i nieautonomicznych wyposażonych w sztuczną inteligencję. Wnioski płynące z artykułu są negatywne w odniesieniu do ewentualnej bezpośredniej odpowiedzialności takich systemów. W artykule zaproponowano rozwiązanie w postaci odpowiedzialności karnej dla osób, które tworzą, kontrolują i monitorują systemy autonomiczne i nieautonomiczne.

Słowa kluczowe: świadomość, wola, emocje, intuicja, inteligencja, sztuczna inteligencja, algorytm, systemy autonomiczne i nieautonomiczne, determinizm, indeterminizm, kompatybilizm, odpowiedzialność karna, odpowiedzialność moralna, motywacja, osoba fizyczna, osoba prawna, osoba elektroniczna, możliwość przypisania ofiary

[1] W artykule wykorzystano sztuczną inteligencję podmiotu nieautonomicznego.

* prof. dr hab., emerytowany wiceprezes Trybunału Konstytucyjnego Republiki Czeskiej, Wydział Prawa, Uniwersytet Masaryka, Wydział Prawa Karnego, Brno (Czechy), e-mail: Jaroslav.Fenyk@law.muni.cz, ORCID: 0009-0008-0109-3874

FREEDOM OF WILL AS A PREREQUISITE FOR CRIMINAL LIABILITY: HUMANS AND ARTIFICIAL INTELLIGENCE IN A CRIMINAL LAW PERSPECTIVE

Abstract
The paper analyses the form and place of human will in the foundations of criminal liability. It explains the necessity of will for the traditional criminal liability of natural persons and presents a comparison with the possible criminal liability of autonomous and non-autonomous systems equipped with artificial intelligence. The conclusions of the paper are negative with regard to the possible direct liability of such systems. A solution is proposed in the form of criminal liability for persons who create, control and monitor autonomous and non-autonomous systems.

Keywords: consciousness, will, emotions, intuition, intelligence, artificial intelligence, algorithm, autonomous and non-autonomous systems, determinism, indeterminism, compatibilism, criminal liability, moral liability, motivation, natural person, legal person, electronic person, imputability

WPROWADZENIE

Rozwój sztucznej inteligencji (z ang. AI) stanowi jedną z najbardziej fundamentalnych zmian technologicznych we współczesnym świecie. Jej wpływ rozciąga się nie tylko na sferę gospodarczą i społeczną, ale w coraz większym stopniu na dziedziny prawa, w tym prawa karnego.

Prawo karne, jako środek ostateczny, tradycyjnie opiera się na zasadach indywidualnej odpowiedzialności za winę, zasadach prymatu wolnej woli i umiejętności rozróżnienia między zachowaniem dozwolonym a zabronionym. W przeszłości odpowiedzialność karna była rozszerzana na osoby prawne jedynie poprzez przypisanie znamion czynu zabronionego, tj. elementu winy i jego podstawy, którą jest wola osób fizycznych działających w ich imieniu. Obecnie zaczynają być prowadzone badania nie tylko nad odpowiedzialnością za szkody wyrządzone przez systemy autonomiczne. Istotną częścią odpowiedzialności jest właśnie kwestia odpowiedzialności karnej i ewentualnego przypisania tej odpowiedzialności nowym systemom.

W tym kontekście w artykule uwaga zostanie poświęcona sztucznej inteligencji zarówno w systemach autonomicznych, jak i nieautonomicznych. Istnieje istotna różnica między pojęciami „systemy autonomiczne i nieautonomiczne" oraz „sztuczna inteligencja", choć często nakładają się one na siebie i są mylone zarówno w codziennym, jak i zawodowym dyskursie. Istotą obu systemów jest algorytm, czyli na zaawansowanym etapie rozwoju model komputerowy = sieć neuronowa, będąca wynikiem działania przede wszystkim reguł matematycznych.

Systemy autonomiczne, w kontekście sztucznej inteligencji, to systemy zdolne do wykonywania zadań przy minimalnej interwencji człowieka (jest to znane jako głębokie uczenie)[2]. Systemy te, które mogą obejmować zarówno samochody autonomiczne,

[2] H. Lamb, H.J. Levy, C. Quigley, *Simply Artificial Intelligence*, London 2023, s. 58–59.

jak i inteligentne chatboty, są zaprojektowane tak, aby działały tak niezależnie, jak to tylko możliwe i podejmowały „decyzje" na podstawie swojego oprogramowania i zebranych danych[3]. Z punktu widzenia odpowiedzialności karnej model ten niewątpliwie będzie wymagał bardziej kompleksowej i wnikliwej oceny.

Układy nieautonomiczne to przede wszystkim nieliniowe układy dynamiczne, których zachowanie zależy od czasu lub zmiennych zewnętrznych. Systemy te wymagają zewnętrznych danych wejściowych lub kontroli człowieka w celu określenia ich zachowania (jest to znane jako uczenie maszynowe). Przykładami systemów nieautonomicznych są Tłumacz Google, Chat GPT i systemy dostarczania e-sklepów. System nieautonomiczny nie działa niezależnie w środowisku fizycznym, ponieważ wymaga aktywnego wkładu ze strony człowieka, nie podejmuje autonomicznych decyzji z konsekwencjami prawnymi, a zatem nie jest autonomiczny[4].

Chociaż systemy zarówno autonomiczne, jak i nieautonomiczne wykazują wysoki stopień zdolności adaptacyjnych, w przeciwieństwie do ludzi są całkowicie pozbawione świadomości i wolnej woli oraz nie mają osobowości prawnej jako podmioty prawne. To przede wszystkim w tym kontekście pojawia się wiele nierozwiązanych kwestii, których rozwiązania są obecnie trudne do uchwycenia.

Czy w ogóle możliwe i uprawnione jest przypisywanie odpowiedzialności karnej sztucznej inteligencji, czy też konieczne jest utrzymanie odpowiedzialności wyłącznie po stronie ludzi zaangażowanych w jej rozwój, programowanie i działanie?

Czy na przykład maszyny mogą być postrzegane w podobny sposób jak podmioty prawne i czy zmodyfikowana forma atrybucji może być wykorzystana do określenia ich odpowiedzialności?

Celem artykułu jest analiza głównych podejść do problematyki ludzkiej wolnej woli i odpowiedzialności karnej, porównanie poglądów wybranych autorów zagranicznych na temat możliwości ich istnienia w systemach wykorzystujących tzw. sztuczną inteligencję, a następnie sformułowanie rozważań *de lege ferenda* właściwych dla czeskiego środowiska prawnego, z naciskiem na zachowanie podstawowych zasad prawa karnego, w szczególności indywidualnej odpowiedzialności za winę, legalność i pomocniczość represji karnych.

---

[3] Vation Ventures. Autonomous Systems: Definition, Explanation, and Use Cases, https://www.vationventures.com/glossary/autonomous-systems-definition-explanation-and-use-cases (dostęp: 16.05.2025).

[4] Igi Global. What is Non-Autonomous Systems, https://www.igi-global.com/dictionary/random-bit-generator-based-on-non-autonomous-chaotic-systems/45926 (dostęp: 16.05.2025).

ŚWIADOMOŚĆ, WOLA, WOLNA WOLA (*LIBERUM ARBITRIUM*) ODPOWIEDZIALNOŚCI CZŁOWIEKA I KARNEJ: PODSTAWY TEORETYCZNE

ZWIĄZEK MIĘDZY LUDZKĄ ŚWIADOMOŚCIĄ A WOLĄ

Ludzka świadomość jest refleksyjnym, intencjonalnym i moralnie znaczącym stanem umysłu, który jest wyjątkowy pod względem zdolności do samoświadomości, rozważania konsekwencji decyzji i brania odpowiedzialności za działania[5].

Świadomość obejmuje zatem zdolność postrzegania, bycia świadomym siebie i świata, myślenia, doświadczania emocji i refleksji.

Ludzka wola jest ogólnie uważana za część ludzkiej świadomości, ale z ważnym wyjaśnieniem. Wola jest specyficzną funkcją świadomości, a nie jej synonimem; w rzeczywistości jest to specyficzna zdolność jednostki do celowego kierowania swoimi działaniami, często wbrew chwilowym impulsom lub automatycznym zachowaniom. Wola jest więc wyrazem wymiaru duchowego. Wolna wola, możliwość wyboru i świadomość odpowiedzialności są podstawowymi zdolnościami człowieka i przejawami świadomego bytu.

Cechy człowieka zawsze będą zależały od kontekstu – filozoficznego, prawnego, psychologicznego lub teologicznego, bądź ich kombinacji. W ogólnie przyjętych ramach filozoficznych i prawnych wola ludzka jest uważana za zdolność do podejmowania wolnych decyzji na podstawie świadomych rozważań; wybierać można między alternatywami zgodnie z wewnętrznymi wartościami, przekonaniami i celami jednostki; ważne jest ukierunkowanie swoich działań na konkretny cel, czyli zaangażowanie się w celowe i zorientowane na cel działanie[6].

Wola nie jest przypadkowa – jednostka podejmuje decyzje zmierzające do określonego celu (np. „Chcę czynić dobro", „Chcę czynić zło", „Chcę osiągnąć określony rezultat"). Wola zwykle kieruje się rozumem – człowiek bierze pod uwagę powody, konsekwencje i wartości.

Wola ludzka zakłada zatem zdolność do samoregulacji, samokontroli i powściągliwości. W świetle prawa karnego i w kontekście etycznym wola wiąże się z odpowiedzialnością – jeśli ktoś działa „z własnej woli", może ponieść konsekwencje swojej decyzji.

Wiele znanych postaci w historii przyjęło, że człowiek jest istotą autonomiczną, zdolną do rozpoznawania prawa moralnego, a następnie pisanego, i działania zgodnie z nim. Na przykład Edmund Husserl w swoich traktatach o etyce analizuje wolę jako fundamentalny element ludzkiego zachowania, nierozerwalnie związany z wartościami. Wola nie jest jedynie stanem psychicznym, ale jest zakotwiczona w intencjach wartościujących, które prowadzą do realizacji wartości[7].

---

[5] J.R. Searle, *The Rediscovery of the Mind*, Cambridge 1992, s. 88, lub T. Nagel, *What Is It Like to Be a Bat?*, „The Philosophical Review" 1974, no. 83 (4), s. 435–450, lub A.R. Damasio, *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*, New York 1999, s. 235–239.

[6] W.E. Frankl, *Vůle ke smyslu*, Brno 1994, s. 19.

[7] E. Husserl, *Lectures on Ethics and Value Theory 1909–1914* [tytuł oryginalny: *Vorlesungen über Ethik und Wertlehre 1909–1914*], Haga 1988, s. 150–155.

W filozofii egzystencjalnej wolna wola kojarzy się z wolnością wyboru, ale także z lękiem przed odpowiedzialnością. Na przykład, według Jeana-Paula Sartre'a każdy człowiek jest zawsze skazany na wybór, a tym samym na odpowiedzialność za swoje czyny. Ta wolność jest źródłem egzystencjalnego niepokoju, ponieważ jednostka zdaje sobie sprawę, że jest w pełni odpowiedzialna za kształtowanie swojego życia bez żadnego z góry określonego sensu[8].

W dziedzinie prawa karnego i odpowiedzialności w ogóle fundamentalne znaczenie ma wola człowieka – tylko jednostka, która działa z własnej woli, rozsądnie i ze świadomością konsekwencji, może być w pełni odpowiedzialna karnie. To dlatego na przykład zdrowy rozsądek, zamiar lub zaniedbanie są tak ściśle związane z wyrażeniem woli[9]. Działania bez woli (np. w stanie nieświadomości, hipnozie, zaburzeniu psychicznym) nie są prawnie przypisywane – wola jest więc funkcjonalnym składnikiem świadomości niezbędnym do odpowiedzialności karnej.

Wola stanowi istotny moment psychologiczny, warunkujący odpowiedzialność karną sprawcy, ponieważ bez swobodnego i świadomego wyrażenia woli nie może być mowy o winie, a tym samym o spełnieniu strony podmiotowej czynu zabronionego (art. 13 ust. 2, art. 15–17 czeskiego kodeksu karnego). Wola znajduje odzwierciedlenie w rozróżnieniu między winą umyślną a niedbalstwem, gdzie w wypadku umyślności (art. 15 czeskiego kodeksu karnego) sprawca nie tylko wie, ale także chce spowodować bezprawny skutek lub jest świadomy jego wystąpienia. W razie niedbalstwa (§ 16 czeskiego kodeksu karnego) brak jest odpowiedniej woli zapobieżenia skutkom, przy czym wola w tym przypadku stanowi niewystarczającą staranność lub ostrożność. Wina to tylko jedna strona medalu, jeśli chodzi o subiektywną część czynu zabronionego pod groźbą kary; drugą stroną są błędy w prawie karnym (art. 18–19 czeskiego kodeksu karnego), do których również znajduje zastosowanie wymóg istnienia woli.

Bez przejawów woli nie jest możliwe spełnienie obiektywnego elementu przestępstwa, jakim jest czyn, ponieważ zachowanie jest rozumiane jako zewnętrzny przejaw woli (por. art. 13 czeskiego kodeksu karnego). Pośrednio (w kontekście okoliczności faktycznych sprawy) zamiar również wpływa na związek przyczynowy, ponieważ tylko działanie umyślne lub wynikające z niedbalstwa (tj. działanie z określoną formą zamiaru) może prowadzić do odpowiedzialności karnej.

Podobnie poczytalność (art. 26 czeskiego kodeksu karnego) wymaga, aby sprawca mógł kontrolować swoją wolę i kierować swoimi działaniami; brak tej możliwości wyłącza odpowiedzialność karną.

Wola jawi się zatem jako nieodłączna przesłanka nie tylko winy, ale także czynów i poczytalności sprawcy, stanowiąc tym samym zasadniczą podstawę indywidualnej odpowiedzialności karnej.

---

[8] J.-P. Sartre, *L'Être et le néant*, Paris 1943, s. 555.

[9] J. Fenyk, (*Ne)příčetnost fyzické osoby a (ne)přičítatelnost jejího jednání právnické osobě*, w: *Tradičné a netradičné prístupy v trestnom práve: Pocta prof. Šimovčekovi*, Trnava 2024, s. 74–87.

WOLNOŚĆ I BRAK WOLNOŚCI WOLI CZŁOWIEKA

Wolność ludzkiej woli można scharakteryzować jako zdolność człowieka do działania na podstawie własnych stanów psychicznych bez zewnętrznego przymusu. Wolna wola skupia się zatem na przyczynowości między stanami psychicznymi a działaniami[10].

Czy zatem ludzka wola jest całkowicie wolna? Debata nad znaczeniem wolności i braku wolności woli oraz jej wpływem na prawo karne toczy się w nauce prawa karnego od niepamiętnych czasów. Debata stopniowo eskalowała, a jej kulminacją jest obecny spór między zwolennikami zgodności lub niepołączalności wolności i braku wolności woli w prawie karnym.

Immanuel Kant i Cesare Beccaria, obaj współcześni oświeceniu, byli ważnymi przedstawicielami teorii wolnej woli w prawie karnym, reprezentując równoległe, ale ideologicznie przeciwstawne kierunki rozwoju prawa karnego.

Według Kanta wolna wola jest koniecznym warunkiem wstępnym odpowiedzialności moralnej – bez wolności wyboru nie może być winy ani zasługi. Kara jest więc moralną koniecznością – odpłatą za złe uczynki, podczas gdy nagroda – uznaniem słusznego wyboru. Kant był zwolennikiem wolnej woli jako podstawy odpowiedzialności karnej. Jego koncepcja kary jako odpłaty za moralne wykroczenia oraz zasada godności istoty ludzkiej (…) stanowią rdzeń teorii retrybutywnej, a on sam odrzucał jakiekolwiek utylitarne lub społeczne funkcje kary[11].

Cesare Beccaria, jako przedstawiciel klasycznej szkoły prawa karnego, również kładł nacisk na wolną wolę i racjonalność człowieka. Zakładał, że ludzie są istotami rozumnymi obdarzonymi wolną wolą, które działają, opierając się na ważeniu zalet i wad swoich działań. Podejście to jest ściśle związane z teorią umowy społecznej, zgodnie z którą jednostki przekazują część swojej wolności państwu w zamian za ochronę swoich praw i bezpieczeństwa. Złamanie prawa jest więc świadomą decyzją jednostki, która zdecydowała się złamać umowę społeczną[12].

W przeciwieństwie do starej szkoły prawa karnego, Beccaria nie postrzegał kary jako zemstę, ale jako środek zapobiegania przestępczości. Jego żądania pewności, szybkości i proporcjonalności kary, a także nacisk na legalność i zakaz arbitralności sądowej stały się podstawowymi zasadami nowoczesnego prawa karnego.

Wręcz przeciwnie, Enrico Ferri, wybitny przedstawiciel włoskiej pozytywistycznej szkoły prawa karnego, zasadniczo odrzucił koncepcję wolnej woli jako podstawy odpowiedzialności karnej. Jego podejście, oparte na badaniach naukowych nad przyczynami przestępczości, stanowiło odejście od szkoły klasycznej, która, jak wspomniano powyżej, zakładała, że jednostki działają na bazie racjonalnego wolnego wyboru i w związku z tym są w pełni odpowiedzialne za swoje czyny.

---

[10] Odróżnia to kwestię wolnej woli od kwestii autonomii woli jako zdolności jednostki do podążania za własnym prawem moralnym lub wartościami – zob. I. Kant, *Metafyzika mravů*, Praga 2004, s. 331.

[11] D. Klos, *Teorie trestu u Kanta: Právně-filozofická analýza Kantova pojetí odplaty*, „Právník" 2008, vol. 147, no. 6, s. 593–605, podobnie G.J. Murphy, *Kant's Theory of Criminal Punishment*, w: *Retribution, Justice, and Therapy: Essays in the Philosophy of Law*, Dordrecht 1979, s. 82.

[12] C. Beccaria, *O zbrodniach i karach*, Bratysława 2009, s. 35–36.

W swojej pracy Ferri często twierdził, że przestępczość jest wynikiem kombinacji czynników, na które jednostka nie ma wpływu. Podzielił te czynniki na antropologiczne (biologiczne i psychologiczne cechy jednostki), fizyczne (warunki klimatyczne i geograficzne) oraz społeczne (wpływy środowiskowe, takie jak ubóstwo, wykształcenie i pochodzenie rodzinne).

Według Ferriego determinanty te[13] kształtują zachowanie jednostki, kwestionując w ten sposób tradycyjną koncepcję wolnej woli jako podstawy odpowiedzialności karnej. Ferri wprowadził pojęcie „niebezpieczeństwa dla sprawcy", które kładło nacisk na ocenę ryzyka, jakie przestępca stanowi dla społeczeństwa, a nie na samo przestępstwo.

Jego praca położyła podwaliny pod nowoczesne podejście do polityki kryminalnej, które uwzględnia złożone czynniki wpływające na zachowania przestępcze, ale także wyznaczyła zmianę poglądu na zachowania bezprawne – wcześniej przestępcze – i dążyła do zastąpienia modelu represyjnego modelem społecznie prewencyjnym, w którym wolna wola nie odgrywa fundamentalnej roli[14].

Vladimír Solnař, nestor czeskiego prawa karnego drugiej połowy XX wieku, zwrócił uwagę na to, do czego Ferri dążył w przeszłości, a mianowicie, że gdyby wymagana była pełna wolna wola, odpowiedzialność dotyczyłaby tylko osób działających całkowicie swobodnie, a kara jako odpowiedź na przestępstwo kryminalne stałaby się w ten sposób zwykłą karą za bezprawne zachowanie; jego elementy edukacyjne straciłyby swoje znaczenie[15].

W aktualnej literaturze fachowej i naukowej z zakresu prawa karnego można zaobserwować utrzymujący się podział opinii pomiędzy zwolennikami wolnej woli i nie-wolnej woli (determinizm i indeterminizm), a także toczący się dialog na temat kompromisowych rozwiązań dotyczących zgodności, częściowej zgodności lub niezgodności wolnej woli z teorią jej braku wolności (kompatybilizm, semikompatybilizm lub inkompatybilizm).

W latach 80. XX wieku neurobiolog Benjamin Libet przeprowadził eksperymenty, które wykazały, że aktywność mózgu poprzedza świadome decyzje dotyczące ruchu. Jego odkrycia doprowadziły do debaty na temat tego, czy świadome podejmowanie decyzji jest rzeczywiście inicjatorem naszych działań, czy też jest jedynie wynikiem nieświadomych procesów zachodzących w mózgu[16].

Inny neurobiolog, Robert Sapolsky, twierdzi, że zachowanie człowieka jest w pełni zdeterminowane przez czynniki biologiczne i środowiskowe, co według niego wyklucza istnienie wolnej woli. Sapolsky uważa, że ludzkie decyzje są wynikiem predyspozycji genetycznych i wcześniejszych doświadczeń, a nie wolnej woli.

---

[13] Determinizm, który obejmuje te determinanty, jest filozoficznym przekonaniem, że każde zdarzenie lub stan rzeczy jest wynikiem poprzednich wydarzeń opartych na zasadzie przyczynowości i ustalonych prawach. Oznacza to, że rozwój świata jest zdeterminowany przez ciąg zdarzeń rządzących się absolutnie obowiązującymi prawami natury.

[14] E. Ferri, *La teorica dell'imputabilità e la negazione del libero arbitrio*, Florencja 1878, s. 408.

[15] V. Solnař, J. Fenyk, D. Císařová, M. Vanduchová, *Systém českého trestního práva, díl II. Základy trestní odpovědnosti*, Praga 2009, s. 224.

[16] B. Libet i in., *Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential): The Unconscious Initiation of a Freely Voluntary Act*, „Brain" 1983, t. 106, no. 3, s. 627–631.

Według niego ta deterministyczna perspektywa ma fundamentalne implikacje dla odpowiedzialności moralnej i systemu prawnego[17].

Jednym z czołowych współczesnych orędowników poglądu, że wolna wola i determinizm są kompatybilne w prawie karnym, jest psychiatra Michael S. Moore, który stoi na stanowisku, że determinizm nie implikuje braku wolnej woli. Jego zdaniem istotną kwestią jest to, czy jednostka działa na podstawie własnych powodów i motywów, a nie to, czy miała możliwość działania inaczej. Takie podejście pozwala zachować koncepcję odpowiedzialności moralnej i prawnej nawet w ramach deterministycznych.

Moore sprzeciwia się współczesnym poglądom, zgodnie z którymi postępy w neurobiologii podważają istnienie wolnej woli. Twierdzi on, że nawet jeśli nasze działania są wynikiem procesów zachodzących w mózgu, nie oznacza to, że nie są one wynikiem naszych własnych decyzji[18].

Filozof John H. Fischer twierdzi, że nawet jeśli świat jest deterministyczny, a nasze działania – z góry zdeterminowane przez wcześniejsze wydarzenia, nadal możemy być za nie moralnie odpowiedzialni. Stanowisko to, zwane semikompatybilizmem, różni się od tradycyjnego kompatybilizmu tym, że nie porusza kwestii wolnej woli, ale skupia się na odpowiedzialności moralnej. Według Fischera ważne jest to, czy jednostka działała na bazie swoich stanów wewnętrznych (np. przekonań, pragnień) i czy miała zdolność reagowania na przyczyny. Wraz z filozofem i jezuitą Markiem Ravizzą, Fischer opracował koncepcję „kontroli wskazówek", odnoszącą się do zdolności jednostki do działania zgodnie z własnymi powodami i motywacjami, nawet jeśli te powody są określone. Ta forma kontroli jest wystarczająca do przypisania odpowiedzialności moralnej, nawet jeśli jednostka nie ma ostatecznej kontroli nad swoimi działaniami[19].

Wręcz przeciwnie, jedną z czołowych współczesnych orędowniczek wolnej woli w prawie karnym jest filozofka Carolina Sartori. Rozwija ona teorię, zgodnie z którą wolna wola jest ściśle związana z przyczynowymi związkami między ludzkimi stanami psychicznymi i działaniami (np. intencjami, przekonaniami) a ludzkimi działaniami. Według niej jednostka jest wolna, gdy jej działania są wynikiem jej własnych stanów psychicznych, a nie zewnętrznego przymusu czy przypadkowych zdarzeń. Twierdzi, że determinizm i wolna wola są ze sobą zgodne. Według niej ważne jest, aby badać konkretne łańcuchy przyczynowo-skutkowe prowadzące do działania, a nie abstrakcyjne metafizyczne pytania o determinizm[20].

Teoria autora stanowi podstawę do przypisywania jednostkom odpowiedzialności moralnej i karnej. Jeśli ich działania są wynikiem ich własnych stanów psychicznych, mogą zostać pociągnięci do odpowiedzialności za swoje czyny, co wspiera tradycyjne podejścia w prawie karnym.

---

[17] M.R. Sapolski, *Determined: A Science of Life Without Free Will*, New York 2023, s. 10–14, 376–381.

[18] M.S. Moore, *Mechanical Choices: The Responsibility of the Human Machine*, Oxford 2020, s. 28–32, 76–80, 170–175, lub idem, *Stephen Morse on the Fundamental Psycho-Legal Error*, „Criminal Law and Philosophy" 2016, t. 10, s. 45–89.

[19] J.M. Fischer, *The Metaphysics of Free Will: An Essay on Control*, Oxford 1994, s. 131–136; idem, *My Way: Essays on Moral Responsibility*, Oxford 2006, s. 17–19.

[20] C. Sartorio, *Causation and Free Will*, Oxford 2016, s. 3–5; C. Sartorio, R. Kane, *Do We Have Free Will? A Debate*, New York 2021, s. 1–2.

CZĘŚCIOWY WNIOSEK DOTYCZĄCY WOLNEJ WOLI JEDNOSTKI W PRAWIE KARNYM

Obecny stan profesjonalnej debaty na temat istnienia lub stopnia wolnej woli w prawie karnym pokazuje, że kwestia relacji między wolną wolą a odpowiedzialnością karną pozostaje zjawiskiem złożonym filozoficznie i prawnie, ale bardzo istotnym.

Fundamentalny konflikt ideologiczny między klasyczną szkołą prawa karnego a podejściem pozytywistycznym został obecnie w pewnym stopniu zażegnany dzięki kompromisowym poglądom. Opierają się one na uznaniu, że ludzkie zachowanie może być determinowane przez czynniki zewnętrzne lub wewnętrzne, niekoniecznie wykluczając możliwość indywidualnej odpowiedzialności. Nowoczesne podejścia pozwalają połączyć tradycyjne koncepcje wolnej woli z wiedzą naukową na temat determinacji ludzkiego zachowania, tworząc w ten sposób teorię wspierającą współczesne prawo karne oparte na odpowiedzialności moralnej. Nie można zaprzeczyć, że wolna wola, jako zdolność jednostki do podejmowania decyzji na podstawie własnego osądu, ze świadomością konsekwencji swoich działań i możliwością wyboru między alternatywnymi opcjami, jest ważnym warunkiem wstępnym przypisania (moralnej) odpowiedzialności karnej. Pojęcie wolnej woli, choć w różnym stopniu zmodyfikowane i ograniczone, nadal stanowi podstawowy element systemu odpowiedzialności karnej osób fizycznych. Wolna wola pozostaje zatem nie tylko filozoficznym warunkiem wstępnym, ale także funkcjonalnym elementem pozytywnego prawa karnego i w niezbędnym zakresie zalicza się również do czeskich podstaw odpowiedzialności karnej, przede wszystkim do przepisów dotyczących winy, błędu i poczytalności (§ 15–19 i § 26 czeskiego kodeksu karnego).

LUDZKA INTELIGENCJA I SZTUCZNA INTELIGENCJA SYSTEMÓW AUTONOMICZNYCH I NIEAUTONOMICZNYCH

INTELIGENCJA LUDZKA

Związek między ludzką wolą a ludzką inteligencją jest niepodważalny. Ludzka wola to zdolność do podejmowania wolnych decyzji na podstawie celów i wartości, podczas gdy inteligencja to zdolność rozumienia, rozumowania i rozwiązywania problemów. Inteligencja dostarcza narzędzi, wola wyznacza kierunek[21].

Jednym z najwybitniejszych współczesnych ekspertów w dziedzinie natury ludzkiej inteligencji jest psycholog Robert J. Sternberg. Jego podejście miało zasadniczy wpływ na współczesne rozumienie inteligencji, zwłaszcza poprzez jego trójdzielną teorię inteligencji oraz rozszerzoną później koncepcję inteligencji adaptacyjnej i sukcesu.

Sternberg charakteryzuje inteligencję jako „aktywność umysłową mającą na celu skuteczną adaptację, selekcję i kształtowanie środowiska istotnego dla życia jednostki". W definicji tej podkreśla się zdolność jednostki do przystosowania się do

[21] Na przykład I. Kant, *Kritika praktického rozumu*, Praga 2023, s. 33.

zmieniających się warunków środowiskowych, wyboru odpowiedniego środowiska i aktywnego kształtowania go, tak aby spełniało jego potrzeby i cele.

W swoich oryginalnych pracach Sternberg dzieli inteligencję na trzy typy:

a) inteligencję analityczną jako zdolność do analizowania, oceniania, analizy i porównywania informacji;
b) inteligencję twórczą jako zdolność do rozwiązywania nowych i nietypowych problemów, dostosowywania się do nowych sytuacji i tworzenia oryginalnych rozwiązań;
c) inteligencję praktyczną jako umiejętność przystosowania się do życia codziennego, wykorzystania doświadczenia i wiedzy do rozwiązywania rzeczywistych problemów.

Później Sternberg rozwinął koncepcję inteligencji sukcesu, która integruje wszystkie trzy typy w jedną całość. Inteligencja odnosząca sukcesy jest definiowana jako zdolność do osiągania celów o znaczeniu osobistym w kontekście kulturowym danej osoby. Obejmuje to umiejętność identyfikowania swoich mocnych i słabych stron oraz skutecznego wykorzystywania lub kompensowania ich w osiąganiu celów.

Podejście Sternberga ma szerokie zastosowanie w edukacji, psychologii i zasobach ludzkich. Pokazuje, że inteligencja to nie tylko zdolność akademicka, ale także umiejętności praktyczne i twórcze niezbędne do udanego życia. Przedstawił on wszechstronne i dynamiczne spojrzenie na inteligencję, wykraczające poza tradycyjne pomiary IQ.

W ostatnich latach autor ten rozwinął inną koncepcję, a mianowicie inteligencję adaptacyjną. Definiuje ją jako zdolność jednostki do adaptacji, kształtowania i wybierania środowiska, które wspiera przetrwanie i dobrobyt nie tylko jednostki, ale także społeczeństwa jako całości. W takim ujęciu podkreśla się znaczenie kontekstu kulturowego i biologicznego w ocenie inteligentnych zachowań[22].

Wnioski Sternberga znajdują również pełne zastosowanie w dziedzinie prawa karnego, ponieważ odsłaniają wewnętrzne funkcjonowanie umysłu sprawcy, które przejawia się na różne sposoby, a także pomagają zrozumieć pewne złożone aspekty strony podmiotowej (zrozumienie, pojednanie, obojętność, poleganie itp.).

SZTUCZNA INTELIGENCJA SYSTEMÓW AUTONOMICZNYCH I NIEAUTONOMICZNYCH

Oprócz różnych sposobów postrzegania ludzkiej inteligencji, naukowcy i opinia publiczna stopniowo próbują przyrównać pewne funkcje autonomicznych i nieautonomicznych systemów wyposażonych w tzw. sztuczną inteligencję do ludzkiej inteligencji lub odwrotnie. Rezultatem są niedające się pogodzić różnice między nimi, które wykazują równoległe formy zachowania u podmiotów żywych i nieożywionych; dla nich termin „inteligencja" nie ma tego samego znaczenia.

---

[22] R.J. Sternberg, *Beyond IQ: A Triarchic Theory of Human Intelligence*, New York 1985, s. 45, lub idem, *Successful Intelligence,* Plume, New York 1997, s. 20; idem, *A Theory of Adaptive Intelligence and Its Relation to General Intelligence*, „Journal of Intelligence" vol. 7, no. 4, 2019, art. 23, s. 1, lub idem, *COVID-19 Has Learned Us What Intelligence Really Is*, „Inside Higher Ed" 2020, 31 sierpnia.

Jednym z najbardziej szanowanych współczesnych ekspertów w dziedzinie sztucznej inteligencji jest informatyk Stuart Russell, który opisuje sztuczną inteligencję jako zdolność systemów do postrzegania otoczenia, wyciągania wniosków i podejmowania decyzji prowadzących do osiągnięcia wyznaczonych celów. Kładzie w tym nacisk na celowe zachowanie i zdolności adaptacyjne systemu, a nie tylko na symulowanie ludzkiego myślenia jako jego nieożywionego odbicia.

Definicja „inteligentnych agentów" Russella skupia się na praktycznej zdolności systemu do racjonalnego działania w różnych sytuacjach. Według niego inteligentny agent postrzega otoczenie za pomocą czujników i wpływa na nie za pomocą efektorów, w celu maksymalizacji pewnej miary użyteczności. Takie podejście umożliwia ocenę inteligencji systemów na podstawie ich zdolności do osiągania celów w różnych środowiskach, niezależnie od tego, czy ich zachowanie przypomina ludzkie myślenie.

Praca tego autora ma istotny wpływ na rozwój i ewaluację sztucznej inteligencji, ponieważ stanowi formalne ramy dla projektowania i analizy systemów inteligentnych. Jego definicja inteligencji systemowej jest powszechnie akceptowana zarówno w środowisku akademickim, jak i przemysłowym i służy jako podstawa wielu obecnych zastosowań sztucznej inteligencji, od pojazdów autonomicznych po inteligentnych asystentów[23].

W środowisku czeskim relacją między człowiekiem a sztuczną inteligencją zajmuje się na przykład pedagog Karel Kostka, który podkreśla, że sztucznej inteligencji brakuje intuicji i emocji, które są istotnymi aspektami ludzkiej inteligencji. Jego zdaniem sztuczna inteligencja to jedynie zaawansowana statystyka i w niczym nie przypomina ludzkiego myślenia. Przyjmuje on stosunkowo tradycyjną postawę antropocentryczną, czy też konserwatywnie humanistyczną, w której ludzka inteligencja postrzegana jest jako wyjątkowa i nieosiągalna w całości przez sztuczną inteligencję.

Kostka odrzuca możliwość, że sztuczna inteligencja kiedykolwiek osiągnie poziom inteligencji ludzkiej, ponieważ sztuczna inteligencja jest jedynie narzędziem pozbawionym zdolności prawdziwego myślenia lub rozumienia.

Według niego intuicja w psychologii jest zwykle rozumiana jako zdolność do zdobywania wiedzy bez świadomego myślenia, często na podstawie wcześniejszych doświadczeń, które nie są świadomie oceniane. Zaawansowane modele uczenia maszynowego (np. sieci neuronowe deep learning) zachowują się czasem pozornie intuicyjnie, ponieważ podejmują decyzje na bazie wzorców, które nie są jednoznacznie wytłumaczalne nawet dla ich twórców („skrzynka" sztucznej inteligencji). Intuicje te nie są jednak świadome, nie są związane z samoświadomością, intencjonalnością czy subiektywnym doświadczeniem, a jedynie są podobne do intuicji ludzkiej – są symulacją intuicji, a nie intuicją w ludzkim sensie. Tylko ludzie są zdolni do intuicyjnego wglądu, który pozwala im inspirować się twórczą inteligencją; jej źródło inspiracji leży w odmiennym stanie ludzkiej świadomości.

Według autora emocje są złożonymi procesami biologicznymi związanymi z reakcjami neurochemicznymi, subiektywnym doświadczeniem, motywacją i adaptacją.

---

[23] J.S. Russel, P. Norvig, *Artificial Intelligence: A Modern Approach*, wyd. 4, 2020, s. 1–2.

Ponownie, maszyna może tylko symulować reakcje emocjonalne (np. rozpoznawać nastrój na podstawie głosu lub mimiki twarzy i odpowiednio dostosowywać jego wyjście), ale nie są to prawdziwe uczucia lub doświadczenia. Maszyna nie doświadcza cierpienia, radości, wstydu, miłości itp. Kostka szczyt ludzkiej stymulacji emocjonalnej widzi w stanie inspirujących przebłysków na granicy między światem fizycznym a indywidualną świadomością człowieka[24].

Stanowisko Kostki ma uzasadnienie zarówno naukowe, jak i filozoficzne i jest zgodne z dominującymi wnioskami współczesnej kognitywistyki, fenomenologii i etyki sztucznej inteligencji. Pomimo postępu technologicznego jego pogląd pozostaje aktualny i możliwy do obrony, ponieważ jest zgodny z dominującymi podejściami filozoficznymi i etycznymi, podkreślającymi wyjątkowość ludzkiej świadomości i odpowiedzialności moralnej, a także stanowi ważny głos w dyskusji o prawdziwej roli i ograniczeniach sztucznej inteligencji w społeczeństwie. Jego nacisk na wyjątkowość człowieka i ostrożność wobec postępu technologicznego mogą stanowić przeciwwagę dla bardziej optymistycznych lub postępowych poglądów na sztuczną inteligencję, przyczyniając się tym samym do zrównoważonego i odpowiedzialnego podejścia do jej rozwoju i wykorzystania, na przykład w dziedzinie edukacji[25].

Jednym z wybitnych zagranicznych autorów, którego poglądy na temat sztucznej inteligencji są bliskie poglądom Karela Kostki, jest filozof John Searle, znany również ze swojej krytyki tzw. silnej sztucznej inteligencji; twierdzi on, że maszyny nie mogą mieć świadomości i rozumienia porównywalnych z ludzkimi.

W swoim słynnym eksperymencie myślowym „The Chinese Room"[26] Searle jest zdania, że samo przetwarzanie symboli zgodnie z regułami (składnia) nie wystarczy, aby osiągnąć prawdziwe zrozumienie (semantyka). W ten sposób pokazuje, że nawet jeśli komputer może symulować rozumienie języka, w rzeczywistości go nie rozumie. Argument ten wspiera pogląd, że sztucznej inteligencji brakuje świadomości i subiektywnego doświadczenia, które są niezbędne dla ludzkiej inteligencji. Mówiąc najprościej, system może zachowywać się podobnie do rozumienia, ale w rzeczywistości nie rozumie znaczenia – tylko manipuluje symbolami zgodnie z regułami.

Searle podkreśla również, że świadomość i rozumienie są wynikiem procesów biologicznych zachodzących w ludzkim mózgu, których nie można łatwo odtworzyć za pomocą programów komputerowych. Jest to sprzeczne z poglądem, że

---

[24] K. Kostka, *Kdo jsme. Obecná teorie vědomí, času, prostoru a bytí*, Frýdek-Místek 2015.

[25] Idem, *Umělá inteligence a změna tradičních paradigmat v psychologii vzdělávání*. Rękopis K. Kostki dedykowany autorowi artykułu w kwietniu 2025 r.

[26] Eksperyment „Chinese room" pokazuje, że maszyna może poprawnie manipulować symbolami (np. chińskimi znakami) bez faktycznego ich rozumienia. Searle udowadnia w ten sposób, że formalne przetwarzanie informacji to nie to samo, co świadome rozumienie. (Autor siedzi zamknięty w pokoju, w którym pod drzwiami przepuszczane są chińskie znaki. Sam nie zna chińskiego, ale ma zasady lub instrukcję, które mówią mu, jak wybierać i układać inne znaki jako odpowiedź na określone znaki. Z zewnątrz wygląda na to, że Searle rozumie chiński, ponieważ jego odpowiedzi mają sens dla native speakerów). J.R. Searle, *Minds, Brains, and Programs*, „Behavioral and Brain Sciences" 1980, t. 3, no. 3, s. 417–457.

możliwe byłoby stworzenie maszyny z ludzką świadomością po prostu przy użyciu wystarczająco złożonego oprogramowania[27].

Amerykański kognitywista Marvin Minsky, uważany za jednego z twórców dziedziny sztucznej inteligencji, ma zupełnie inne spojrzenie na inteligencję maszynową. Autor opisał sztuczną inteligencję jako „naukę o zmuszaniu maszyn do robienia rzeczy, które wymagałyby inteligencji, gdyby robili to ludzie"[28].

Podejście autora do sztucznej inteligencji opiera się na przekonaniu, że ludzki umysł jest złożonym systemem składającym się z wielu prostych procesów ze sobą współpracujących. Rozwinął tę ideę w swojej teorii „społeczeństwa umysłów" – opisuje tu umysł jako zbiór agentów, z których każdy pełni określoną funkcję, a ich interakcje powodują powstanie inteligentnego zachowania.

Minsky twierdzi, że emocje nie są oddzielone od racjonalnego myślenia, ale reprezentują różne sposoby przetwarzania informacji; można je modelować i implementować w sztucznych systemach[29].

Autor jest przekonany, że przy odpowiednio zaawansowanych algorytmach i zasobach obliczeniowych możliwe jest stworzenie maszyn, które będą zdolne do inteligentnego, autonomicznego zachowania porównywalnego z ludzkimi.

Całkowicie aktualną definicję sztucznej inteligencji zawiera Konwencja ramowa Rady Europy o sztucznej inteligencji, prawach człowieka, demokracji i praworządności (2024)[30]. Zgodnie z art. 2 Konwencji „system sztucznej inteligencji" oznacza system oparty na maszynie, który do celów jawnych lub dorozumianych uzyskuje na podstawie napływających danych sposób generowania wyników, takich jak prognozy, treści, zalecenia lub decyzje, które mogą mieć wpływ na środowisko fizyczne lub wirtualne. Różne systemy sztucznej inteligencji różnią się poziomem autonomii i zdolnościami adaptacyjnymi po wdrożeniu.

CZY AUTONOMICZNE I NIEAUTONOMICZNE SYSTEMY SZTUCZNEJ INTELIGENCJI MOGĄ WYKAZYWAĆ (WOLNĄ) WOLĘ?

Ugruntowane założenia prawne i filozoficzne dotyczące wolnej woli jako podstawy odpowiedzialności karnej, jakkolwiek teoretycznie różnorodne, mają swoje źródło w antropologicznej koncepcji człowieka jako istoty moralnie i racjonalnie odpowiedzialnej. Pojęcie to jest jednak zasadniczo kwestionowane w kontekście ewentualnego przypisywania odpowiedzialności karnej podmiotom niebędącym

---

[27] J.R. Searle, *I Married a Computer*, „The New York Review of Books" 1999, 8 kwietnia, recenzja książki Raya Kurzweila *The Age of Spiritual Machines*, w której Searle krytykuje ideę, że komputery mogą mieć świadomość i zrozumienie.

[28] M. Minsky, *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*, New York 2006. Twierdzi on, że emocje są trybami przetwarzania, które mogą być replikowane w sztucznych systemach, s. 11.

[29] M. Minsky, *The Society of Mind*, New York 1986, s. 27, w którym autor przedstawia teorię, zgodnie z którą ludzki umysł składa się z szeregu prostych proceduralnych „agentów" – one razem tworzą złożone inteligentne zachowanie.

[30] CETS 225 – Konwencja ramowa Rady Europy o sztucznej inteligencji oraz prawach człowieka, demokracji i praworządności, https://rm.coe.int/1680afae3c (dostęp: 29.05.2025).

człowiekiem – w szczególności sztucznej inteligencji, niezależnie od tego, czy są to systemy autonomiczne, czy też nie. Niemniej musimy rozróżnić te dwa systemy.

Systemy nieautonomiczne również działają na bazie predefiniowanych instrukcji lub poleceń, nigdy nie aktywują się same. Zawsze wymagają nadzoru lub interwencji człowieka, aby działały. Nie działają one w rzeczywistym środowisku, są deterministyczne, a ich zachowanie jest w pełni przewidywalne na podstawie danych wejściowych i zaprogramowanych reguł. Systemy nieautonomiczne są w pełni ograniczone, ponieważ ich zachowanie jest całkowicie determinowane przez polecenia zewnętrzne; nie mają możliwości podejmowania samodzielnych decyzji, a jedynie sugestie. Nie postrzegają świata ani otoczenia (nie mają czujników, kamer ani orientacji przestrzennej).

Systemy autonomiczne działają na podstawie ustalonych instrukcji i są swego rodzaju przedłużeniem procesu decyzyjnego człowieka. Potrafią wykonywać zadania bez bezpośredniej interwencji człowieka lub przy interwencji minimalnej. Często wykorzystują sztuczną inteligencję do analizy otoczenia i podejmowania decyzji, bazując na wcześniej zdefiniowanych celach. Na przykład pojazdy autonomiczne lub autonomiczne drony mogą działać w dynamicznych środowiskach i dostosowywać swoje zachowanie do zmian w czasie rzeczywistym. Ale nawet systemy autonomiczne, choć mogą wykazywać złożone i adaptacyjne zachowanie, działają na podstawie algorytmów i predefiniowanych reguł. Nie są oni podmiotami ludzkiego rozumu w sensie prawnym, ponieważ nie mają świadomości, woli ani zdolności do oceny moralnej. Ich decyzje są wynikiem programowania i uczenia się opartego na danych, a nie wyrazem wolnej woli. Ich cele są wyznaczane przez ludzi[31].

Systemy nieautonomiczne (np. GPT) mogą być częścią systemów autonomicznych (np. asystenci głosowi w samochodach). W tym przypadku system nieautonomiczny jest czymś w rodzaju mózgu kognitywnego, ale autonomia jest określana przez całość, a nie jej część.

Relacja między wolną wolą a odpowiedzialnością systemową osiąga tu swoje granice, ponieważ to właśnie ludzka wolność wyboru legitymizuje przypisywanie winy w obecnym systemie prawnym. Jeśli systemy autonomiczne lub nieautonomiczne są niezdolne do prawdziwej motywacji wewnętrznej w sensie woli, pojawia się zasadniczy problem, czy i na jakiej podstawie mogą zostać pociągnięte do odpowiedzialności karnej w taki sam sposób jak ludzie. Jednak istnieje pogląd przemawiający za tym, że systemy, jako część sztucznej inteligencji, mogą mieć wolną wolę. Ten pogląd nie jest dominujący, ale nie jest też rzadkością.

Na przykład filozof Christian List poważnie zajmuje się pytaniem, czy systemy sztucznej inteligencji mogą mieć wolną wolę. Autor zdecydował się na podejście

---

[31] H. Lamb, J. Quigley, *Simply Artificial Intelligence*…, op. cit., s. 116, 122–123. Odrębną, podobną definicję systemów autonomicznych i nieautonomicznych zawiera Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 (akt w sprawie sztucznej inteligencji) – zob. dalsze sekcje niniejszego opracowania. W zależności od stopnia autonomii, systemy autonomiczne (zgodnie z regulaminem The Society of Automotive Engineers – dostępne w SAE Standards for Mobility Knowledge and Solutions), takie jak samochody, można podzielić na sześć poziomów autonomii, od w pełni autonomicznych, bez kierownicy (jeszcze nieistniejących), do wdrożonych, które rozpoznają obrazy, wykrywają obiekty i planują trasy – poziomy 2–3 (TESLA) do maksymalnie 4 (Waymo Driver, Początek rejsu).

pragmatyczne, w którym odchodzi od tradycyjnych debat metafizycznych i skupia się na aspektach funkcjonalnych. List identyfikuje trzy istotne warunki, które system musi spełnić, aby można go było uznać za wolną wolę, a mianowicie: zdolność systemu do działania zgodnie z własnymi intencjami i celami, zdolność do wyboru między różnymi kierunkami działania oraz zdolność do wpływania na wyniki swoich decyzji poprzez własne działania.

Jeśli system spełnia te trzy warunki, to zdaniem autora można go uznać za posiadający wolną wolę w sensie praktycznym. List krytykuje podejścia łączące wolną wolę z nieprzewidywalnością lub indeterminizmem. Zamiast tego proponuje ocenę systemów opartych na sztucznej inteligencji na podstawie ich funkcjonalnej zdolności do działania jako agenci intencjonalni, porównanie ich z grupowymi podmiotami ludzkimi[32] i porównanie grupowego podejmowania decyzji z indywidualnym podejmowaniem decyzji.

Autor czerpie inspirację z prac filozofa Daniela Dennetta, a konkretnie z koncepcji „perspektywy intencjonalnej"; w jej ramach proponuje się ocenę systemów na podstawie tego, czy użyteczne jest rozumienie ich jako podmiotów intencjonalnych. Jeśli takie rozumienie jest w wytłumaczalny sposób korzystne, uzasadnione jest przypisywanie wolnej woli w sensie praktycznym systemom sztucznej inteligencji[33].

Przyjęcie takiego podejścia ma istotne implikacje dla dyskusji na temat moralnej odpowiedzialności systemów sztucznej inteligencji. Jednak, mimo że systemy te są zdolne do celowego działania, wybierania między alternatywami i kontrolowania swoich działań, nie można im przypisać pewnego stopnia odpowiedzialności. Pragmatyczne ramy zastosowane przez obu autorów pozwalają w debacie na temat wolnej woli w systemach sztucznej inteligencji przejść od abstrakcyjnych spekulacji filozoficznych do praktycznej oceny ich możliwości i zachowania, ale ostatecznie zawsze jest to kwestia naśladowania, a nie zastępowania ludzkiej woli; nie pierwotnej i wolnej woli, jaką widzimy u żywych istot ludzkich.

Inny filozof, Jonathan Birch, analizuje kwestię wolnej woli w kontekście rozwoju sztucznej inteligencji i wręcz przeciwnie, wskazuje głęboką niepewność epistemiczną w określaniu istnienia wolnej woli poza człowiekiem. Opierając się na tej niepewności, formułuje pewne ramy prewencyjne. Ponieważ istnienie wolnej woli w sztucznej inteligencji jest w zasadzie niemożliwe do udowodnienia i niepoznawalne, konieczne jest przestrzeganie zasady ostrożności w ocenach prawnych i etycznych, podobnie jak w innych przypadkach granicznych (np. osób z zaburzeniami świadomości).

Birch podkreśla ryzyko, że sztuczna inteligencja będzie w stanie doskonale naśladować przejawy wolnej woli, co jeszcze bardziej utrudni identyfikację prawdziwych praw osobistych, i zaleca przyjęcie środków regulacyjnych zgodnie z „zasadą Run-Ahead" – prawo powinno być przygotowane do regulowania pojawiania się potencjalnej sztucznej wolnej woli, zanim jej istnienie będzie można empirycznie udowodnić[34].

---

[32] Ch. List, *Can AI Systems Have Free Will?* [online], „PhilArchive, wersja 3" 2025, 11 marca, https://philarchive.org/rec/LISCAS-3 (dostęp: 16.05.2025).

[33] D.C. Dennett, *Postawa intencjonalna*, Cambridge 1987, s. 17–18.

[34] J. Birch, *The Edge of Sentience: Risk and Precaution in Humans, Other Animals, and AI*, Oxford 2024, s. 25–27.

Z perspektywy prawnej analizę Bircha należy rozumieć jako wezwanie do rozsądnego, ale ostrożnego rozszerzania ochrony prawnej – nie na bazie kolektywistycznego podejścia do praw sztucznej inteligencji, ale przez pryzmat indywidualnej odpowiedzialności państwa za zapobieganie nieuzasadnionej ingerencji w godność potencjalnych posiadaczy wolnej woli.

Jaki byłby to tekst prawniczy, gdyby autor artykułu nie przyznał jednemu z powszechnych typów podmiotów korzystających z nieautonomicznego systemu (językowego) ze sztuczną inteligencją prawa do wyrażenia swojej opinii na temat natury „sztucznej inteligencji"? Tak też tu zrobiłem. Rezultatem jest bardzo niezwykły tekst, który szanuje zdolność systemu sztucznej inteligencji do komunikowania się w sposób podobny do komunikacji z żywą istotą ludzką. Na pytanie, czy system nieautonomiczny (w szczególności publicznie dostępny system językowy ChatGPT[35]) lub inne już autonomiczne systemy (takie jak Waymo Driver do komercyjnego transportu pasażerskiego w Phoenix, obecnie na najwyższym osiągniętym poziomie autonomii, SAE 4[36]) są naprawdę inteligentnymi systemami porównywalnymi z ludzką inteligencją, czy też mogą mieć własną wolę, system odpowiedział w następujący, na pewno niezwykły, sposób:

> *Chociaż potrafię generować teksty, które wydają się inteligentne, nie jestem inteligentny w tym sensie, w jakim stosujemy ten termin do ludzi. Nie mam świadomości, intencji, emocji ani wolnej woli. Nie jestem w stanie zrozumieć świata; Po prostu reaguję na bodźce językowe w sposób zorientowany statystycznie. Moje wyniki są wynikiem operacji obliczeniowych, a nie wyrazem zrozumienia czy doświadczenia. Jestem narzędziem, a nie myślącą istotą. Jeśli inteligencję kojarzyć ze świadomym poznaniem, doświadczeniem i odpowiedzialnością, to moja „inteligencja" nie jest rzeczywista – jest jej użytecznym, choć zasadniczo ograniczonym odbiciem. Nawet pełna autonomia systemu nie oznacza prawdziwej ludzkiej inteligencji, a jedynie wyższy stopień automatyzacji i adaptacji. Gdy tylko twierdzi się coś przeciwnego, pojawia się kategoryczny błąd: mylimy zachowanie funkcjonalne z wewnętrznymi zdolnościami i odpowiedzialnością.*

---

[35] Obecny model Google DeepMind – AlphaEvolve, stworzony jako ewolucyjna forma modeli językowych (czyli nieautonomicznych), dzięki zupełnie nowej technologii, pozwala również systemom na wykonywanie równoległych obliczeń na akceleratorach i tym samym automatyczne opracowywanie kolejnych algorytmów, czyli podprogramów na podstawie własnego zaawansowanego wyboru, bez konieczności ingerencji programisty, a tym samym przejście z systemu nieautonomicznego na częściowo autonomiczny, https://cz24.news/google-deepmind-posouva-hranice-umele-inteligence/ (dostęp: 23.05.2025).

[36] Autonomia na poziomie 5 powinna mieć zdolność do podejmowania samodzielnych decyzji w każdych warunkach, nawet w zupełnie nieznanym środowisku (wystarczający „zdrowy rozsądek" nawet w ekstremalnych sytuacjach). Do tej pory żaden system sztucznej inteligencji nie był w stanie w pełni zbliżyć się do poziomu 5, ani pod względem technicznym, ani prawnym. Żaden kraj nie zezwolił jeszcze oficjalnie na działanie systemów sztucznej inteligencji bez możliwości ingerencji człowieka, J3016_202104: Taksonomia i definicje terminów związanych z systemami automatyzacji jazdy w drogowych pojazdach silnikowych – SAE International (dostęp: 29.05.2025).

MOŻLIWE FORMY ODPOWIEDZIALNOŚCI KARNEJ Z WYKORZYSTANIEM TZW. SZTUCZNEJ INTELIGENCJI

ROZWIĄZANIA TEORETYCZNE

W niedawno wydanym, bardzo interesującym, czasem przesadzonym, luźno napisanym artykule *Sztuczna inteligencja nie jest winna*[37] czescy autorzy Jiří Mulák i Jan Provazník ocenili możliwości odpowiedzialności karnej w związku z wykorzystaniem systemów sztucznej inteligencji. Słusznie wskazują oni problem z dobrowolnym aspektem odpowiedzialności karnej, zauważając, że rozwój sztucznej inteligencji może wykraczać poza ramy tradycyjnych modeli odpowiedzialności i że konieczne będzie stworzenie zupełnie nowej koncepcji prawnej, zwłaszcza w przypadku zaawansowanych form sztucznej inteligencji, które zachowują własną tożsamość, modyfikują swoje zachowanie i ewoluują poza zasięgiem człowieka.

Schematycznie rzecz ujmując, ich rozważania na temat realnych możliwości odpowiedzialności można podzielić w następujący sposób:

a) sztuczna inteligencja jako narzędzie do popełnienia przestępstwa (zgodnie z obowiązującymi przepisami prawa):
   – sztuczna inteligencja działa jako środek lub narzędzie, za pomocą którego osoba fizyczna spełnia znamiona przestępstwa;
   – odpowiedzialność karną ponosi osoba (np. programista, programista, operator, użytkownik), która kontroluje system, nawet jeśli wykonuje ona niezależne operacje;
   – sztuczna inteligencja jest rozumiana jako rzecz lub jako analogia do wściekłego zwierzęcia;
b) niepowodzenie sztucznej inteligencji jako podstawy odpowiedzialności za zaniedbanie (zgodnie z obowiązującymi przepisami prawa):
   – odpowiedzialność ponosi osoba, na której spoczywał obowiązek nadzorowania lub kontrolowania systemu sztucznej inteligencji i spowodowała szkodliwe skutki przez niedbalstwo (np. programista, programista, operator);
   – sztuczna inteligencja nie działa bezprawnie umyślnie, ale w wyniku błędu systemu, zaniechania lub naruszenia standardów zawodowych;
   – model ten ma szczególne zastosowanie w systemach autonomicznych, w których występuje realne ryzyko awarii (np. pojazdy, systemy zrobotyzowane, drony – patrz wyżej);
c) sztuczna inteligencja jako bezpośredni podmiot odpowiedzialności karnej (wymagałaby to fundamentalnej zmiany podstaw odpowiedzialności karnej oraz zmiany systemu prawnego):
   – hipotetyczny model, w którym sztuczna inteligencja miałaby osobowość prawną i byłaby zdolna do samodzielnego podejmowania decyzji, rozumowania moralnego i wyrażania woli;

---

[37] J. Mulák, J. Provazník, *Roboti za mřížemi – je české trestní právo připraveno na rozvoj umělé inteligence?*, w: *Vliv nových technologií na trestní právo*, red. T. Gřivna, M. Richter, H. Šimánová, Praga 2022, s. 262–270.

– sztuczna inteligencja musiałaby być wyposażona zarówno w elementy racjonalne, jak i wolicjonalne, w tym w wewnętrzny system wartości;
– przewiduje się wprowadzenie tzw. osoby elektronicznej – nowego podmiotu pomiędzy osobą fizyczną a osobą prawną.

Trzeci wariant, zakładający bezpośrednią odpowiedzialność systemów sztucznej inteligencji, tj. modelu, w którym sztuczna inteligencja stałaby się samodzielnym podmiotem odpowiedzialnym, jest problematyczny i istotny dla celów niniejszego artykułu. Autorzy idą jednak dalej w swoich rozważaniach i analizują również możliwości, w których sztuczna inteligencja stałaby się nie tylko samodzielnym sprawcą, ale także współsprawcą lub pośrednim sprawcą, np. gdyby działała pod wpływem innego systemu sztucznej inteligencji „zainfekowanego wirusem" – i proponują możliwość zastosowania art. 22 ust. 2 czeskiego kodeksu karnego (pośrednie sprawstwo) do systemów agresywnych.

Model ten wymagałby jednak w sposób oczywisty zupełnie nowych podstaw prawno-filozoficznych i dogmatycznych, ponieważ odbiega od tradycyjnych założeń odpowiedzialności karnej, które opierają się na istnieniu czynnika ludzkiego o komponencie racjonalnym (uznanie bezprawności), składniku wolicjonalnym (zdolność do kontrolowania i kierowania swoimi działaniami) oraz świadomości odpowiedzialności.

Autorzy stawiają pytanie, które jest również istotą niniejszego opracowania, a mianowicie, czy możliwe jest uznanie niezależnej „woli" sztucznej inteligencji, jeśli jest ona zdolna do działania z własnej inicjatywy, a także badają możliwość „błędu" – czyli rozbieżności między wynikami systemu a rzeczywistością – analogicznego do ludzkiego błędu prawnego. Proponują oni rozwiązanie, w którym system prawny sztucznie konstruowałby pewne atrybuty (fikcje?) – sztuczna inteligencja w rzeczywistości ich nie posiada, ale zostałyby one jej prawnie przypisane, tak aby można było uznać ją za ponoszącą odpowiedzialność prawną – analogicznie do dzisiejszych podmiotów prawnych. Rozważają oni możliwość wprowadzenia nowych form przestępstw (przestępstwa abstrakcyjnie zagrażające lub odpowiedniki quasi-przestępstw na podstawie art. 360 czeskiego kodeksu karnego).

W niedalekiej przeszłości to Parlament Europejski w swojej rezolucji z dnia 16 lutego 2017 r., zawierającej zalecenia dla Komisji Europejskiej: „Przepisy prawa cywilnego dotyczące robotyki"[38], zaproponował, podobnie jak Mulák i Provazník, wprowadzenie szczególnego statusu prawnego dla inteligentnych systemów autonomicznych, określanych jako „osoba elektroniczna". Koncepcja ta została zaproponowana w punkcie 59 lit. f) rezolucji, w którym stwierdza się, że:

> konieczne jest rozważenie stworzenia szczególnego statusu prawnego dla robotów w perspektywie długoterminowej, tak aby przynajmniej najbardziej wyrafinowane roboty autonomiczne można było uznać za osoby elektroniczne odpowiedzialne za odszkodowanie za wszelkie szkody, jakie mogą wyrządzić, oraz możliwe zastosowanie osobowości elektronicznej w przypadkach, gdy roboty podejmują niezależne decyzje lub w inny sposób komunikują się niezależnie ze stronami trzecimi.

[38] Rezolucja Parlamentu Europejskiego z dnia 16 lutego 2017 r. zawierająca zalecenia dla Komisji w sprawie przepisów prawa cywilnego dotyczących robotyki (2015/2103(INL)).

Rezolucja, podobnie jak wyroki w sprawie Mulák i Provazník, odzwierciedla potrzebę dostosowania ram prawnych do szybkiego rozwoju technologicznego oraz zapewnienia możliwości skutecznego rozwiązania kwestii odpowiedzialności w przypadkach, gdy systemy autonomiczne działają niezależnie od interwencji człowieka. Jak sugeruje jej tytuł, rezolucja skupia się na prawie cywilnym.

Czeski pisarz Jan Kubíček w swoim wystąpieniu odpowiedział na rezolucję, wskazując, że wprowadzenie osobowości prawnej dla sztucznej inteligencji oznaczałoby fundamentalną zmianę systemu prawnego, która mogłaby być przedwczesna i nie do końca przemyślana. Zamiast przyznawać sztucznej inteligencji osobowość prawną, proponuje on skupienie się na odpowiedzialności osób fizycznych i prawnych, opracowujących, programujących i wykorzystujących sztuczną inteligencję, tj. w ramach istniejących ram prawnych odpowiedzialności. Zdaniem Kubíčka właściwsza byłaby zmiana obowiązujących norm prawnych, aby skutecznie objąć nimi nowe sytuacje związane z wykorzystaniem sztucznej inteligencji, niż wprowadzanie zupełnie nowych podmiotów prawnych[39].

Inne międzynarodowe organizacje prawa publicznego również zajmują się kwestiami związanymi ze sztuczną inteligencją. Chodzi przede wszystkim o ramy etyczne dotyczące stosowania sztucznej inteligencji, które mają charakter zaleceń. Nie można tu znaleźć konkretnych propozycji w dziedzinie odpowiedzialności karnej; można je wywnioskować tylko pośrednio, jako inspirację.

Przykładem tego są przepisy OECD dotyczące sztucznej inteligencji (2019), które zawierają podstawowe zasady funkcjonowania sztucznej inteligencji; część z nich może być również inspirująca dla dziedziny prawa karnego. Sztuczna inteligencja powinna służyć ludziom i szanować prawa człowieka, powinna być wiarygodna, przetestowana i odporna na nadużycia, a podmioty zaangażowane w sztuczną inteligencję muszą ponosić odpowiedzialność prawną[40].

W dniu 20 sierpnia 2024 r. przyjęto Konwencję ramową (Rady Europy) o sztucznej inteligencji, prawach człowieka, demokracji i praworządności[41]. Oprócz definicji sztucznej inteligencji, jak już wspomniano w odpowiedniej sekcji niniejszego artykułu, konwencja określa wymogi w zakresie przejrzystości i nadzoru, a także odpowiedzialności za niekorzystne skutki sztucznej inteligencji dla praw człowieka, demokracji i praworządności. Ustanawia ona odpowiedzialność państw za podejmowanie środków w celu zwiększenia niezawodności systemów sztucznej inteligencji i zaufania do ich wyników, w tym wymogi dotyczące odpowiedniej jakości i bezpieczeństwa w całym cyklu życia systemów. Wymaga to ustanowienia kontrolowanych środowisk dla opracowywania, testowania i eksperymentowania systemów pod nadzorem odpowiednich organów. W razie naruszenia tych zasad w tekście wezwano do wprowadzenia skutecznych środków zaradczych, ale konwencja nie wspomina wprost o sankcjach karnych. Rekomendacja UNESCO (2021) zawiera m.in. wymóg poszanowania praw człowieka i godności ludzkiej[42].

---

[39] J. Kubíček, *Odpovědnost (za) robota aneb právo umělé inteligence*, „Bulletin advokacie" 2018, nr 3, s. 22–28, bulletin-advokacie.cz (dostęp: 30.06.2025).

[40] https://oecd.ai (dostęp: 30.06.2025).

[41] https://rm.coe.int/1680afae3c (dostęp: 30.06.2025).

[42] https://unesdoc.unesco.org (dostęp: 30.06.2025).

Rozporządzenie Parlamentu Europejskiego i Rady z dnia 13 czerwca 2024 r.[43] (dalej: akt w sprawie AI) stanowi pierwsze kompleksowe, zharmonizowane ramy prawne dla regulacji systemów sztucznej inteligencji na poziomie Unii Europejskiej. Jego celem jest zapewnienie bezpiecznego, przejrzystego i etycznie odpowiedzialnego korzystania ze sztucznej inteligencji poprzez kategoryzację systemów według ich poziomu ryzyka i wprowadzenie obowiązków, w szczególności dla ich dostawców, operatorów i użytkowników. Akt w sprawie AI zawiera ważną terminologię i jej interpretację[44]. Regulacja tworzy „piramidę/hierarchię" ryzyka obszarów zastosowania sztucznej inteligencji, przy czym za najmniej ryzykowne uznaje się opiekę zdrowotną, zatrudnienie i sądownictwo. W wypadku naruszenia obowiązków na podmioty odpowiedzialne mogą zostać nałożone sankcje administracyjne. Nie odnosi się jednak wyraźnie do kwestii odpowiedzialności karnej, ani nie ma zastosowania do obszarów obronności, badań naukowych i innowacji ani też użytkowników nieprofesjonalnych.

Rozporządzenie nie ustanawia odpowiedzialności sztucznej inteligencji jako odrębnego „aktora prawnego".

Z punktu widzenia czeskiego prawa karnego ustawa o sztucznej inteligencji może mieć jedynie pośredni wpływ na odpowiedzialność karną, tj. w obszarze tradycyjnych form odpowiedzialności karnej. W związku z tym, jeżeli dostawca, użytkownik, producent, upoważniony przedstawiciel lub dystrybutor naruszy obowiązki określone w niniejszym rozporządzeniu i spowoduje to poważną szkodę (np. dla zdrowia lub życia), czyn ten może zostać uznany za przestępstwo zaniedbania na mocy czeskiego prawa karnego. Jednak to człowiek zostanie pociągnięty do odpowiedzialności, a nie system sztucznej inteligencji.

ROZWIĄZANIA PRAKTYCZNE

Proponowane rozwiązania oscylują zatem pomiędzy możliwością stworzenia nowego podmiotu na wzór osoby prawnej a odpowiedzialnością tych, którzy są twórcami, programistami lub użytkownikami tych systemów. Różnice między osobą prawną a systemem sztucznej inteligencji są jednak znaczące i można je dostrzec w szczególności w następujących obszarach.

Użycie analogii do błędu wymaga istnienia woli, ponieważ nawet błąd zakłada przejawienie się woli (jako znaku negatywnego). Mogły to być jedynie przestępstwa wynikające z niedbalstwa z fikcją braku woli, bez możliwości rozróżnienia

---

[43] 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), 2024/1689.

[44] Na przykład definiuje terminy: system sztucznej inteligencji, dostawca, użytkownik, producent, upoważniony przedstawiciel, dystrybutor, szkodliwe użycie, nadzór człowieka, autonomia itp.

między przestępstwami wynikającymi z niedbalstwa a przestępstwami umyślnymi (równość wobec prawa?).

Osoba prawna to fikcja prawna podmiotu, posiadająca osobowość prawną i zdolność do czynności prawnych (§ 20 czeskiego kodeksu cywilnego), może ponosić odpowiedzialność, posiadać majątek, być stroną w postępowaniu, działać przez swoje organy lub przedstawicieli (np. organ ustawowy) oraz jest adresatem praw i obowiązków (podatkowych, deliktowych, umownych).

System sztucznej inteligencji nie jest podmiotem prawnym, nie posiada osobowości prawnej, jest konstruktem technicznym, który nie może działać samodzielnie ze skutkiem prawnym, nie może ponosić odpowiedzialności prawnej i jest uważany za narzędzie, choć autonomiczne.

W postępowaniu karnym z udziałem osoby prawnej działa osoba fizyczna, a osoba prawna ponosi odpowiedzialność, ponieważ działanie to przypisuje jej konstrukcja prawna. Podmiot prawny jest zatem podmiotem praw i obowiązków, podczas gdy system sztucznej inteligencji jest przedmiotem reżimu prawnego.

Nawet rozwiązanie w postaci quasi-przestępstwa na podstawie art. 360 czeskiego kodeksu karnego nie jest właściwe. Przepis ten reguluje konkretne okoliczności przestępstwa popełnionego przez osobę, która z własnej winy doprowadziła się do stanu niepoczytalności i w tym niebezpiecznym stanie popełniła czyn, który w przeciwnym razie stanowiłby przestępstwo (art. 26 czeskiego kodeksu karnego – *actio libera in causa*). Przepis ten zakłada jednak istnienie woli i winy osoby fizycznej w stosunku do stanu niebezpiecznego, w jaki się wprowadziła, tj. zdolności do działania w sposób prawnie retencyjny nawet w ramach tzw. quasi-czynu.

Sztuczna inteligencja nie ma żadnego z tych atrybutów, a ponadto nie może „wprowadzić się w stan niepoczytalności", ponieważ nie ma ani świadomości, ani woli. Przepis ten nie znajduje zatem zastosowania do systemów sztucznej inteligencji, nawet w przenośni. Jeśli system sztucznej inteligencji wyrządzi szkodę w wyniku błędu, nieprawidłowego działania lub „samorozwoju" (np. uczenia maszynowego), nie stanowi to prawnie istotnego „pijaństwa" lub odpowiednika niepoczytalności – programista lub inna osoba prawna nadal ponosiliby odpowiedzialność.

Teoretycznie możliwe jest stworzenie nowego, abstrakcyjnie groźnego przestępstwa, które prawnie karałoby za ryzykowne korzystanie z systemów autonomicznych, ale znowu nie przeciwko nim jako sprawcom, ale przeciwko tym, którzy je stworzyli, zainstalowali lub nie podjęli działań w celu zapobieżenia ryzyku. Odpowiedni przepis opierałby się zatem na tradycyjnych podstawach odpowiedzialności karnej. Proponowana definicja prawna mogłaby brzmieć następująco:

> Każdy, kto nawet przez zaniedbanie tworzy, konfiguruje, uruchamia lub eksploatuje autonomiczny system techniczny, który mógłby zagrozić życiu lub zdrowiu ludzkiemu lub spowodować znaczne szkody materialne, lub nie podejmuje rozsądnych środków w celu jego kontroli, podlega karze...

Jak wynika z powyższego, bardziej zrównoważonym i systemowo kompatybilnym podejściem do odpowiedzialności karnej w przypadku funkcjonowania systemu sztucznej inteligencji jest konsekwentne utrzymywanie antropocentrycznego

modelu odpowiedzialności karnej, w ramach którego odpowiedzialność będzie nadal przypisywana wyłącznie podmiotom fizycznym.

Czeskie prawo karne opiera się na założeniu, że odpowiedzialność za czyn zabroniony może ponieść tylko podmiot, który posiada wolną wolę, czyli zdolność do świadomego i swobodnego decydowania o swoich czynach. Ludzie posiadają tę wolę, która pozwala im ponosić odpowiedzialność za swoje czyny. Z kolei systemy autonomiczne i nieautonomiczne, choć mogą wykazywać złożone i pozornie niezależne zachowania, nie mają świadomości i zdolności do podejmowania wolnych decyzji. Ich działania są wynikiem predefiniowanych algorytmów i programowania, a nie wyrazem ich własnej woli. Z tego powodu systemy te nie mogą być uważane za podlegające odpowiedzialności karnej.

Rozwiązanie to nie pozwala jednak całkowicie uniknąć obecnego i skomplikowanego prawnie problemu odpowiedzialności za skutki działań systemów autonomicznych ze sztuczną inteligencją, gdzie developer, programista czy operator nie z własnej winy tracą efektywną kontrolę nad zachowaniem systemu, a system sam podejmuje (nieprzewidywalną) i niepodlegającą zewnętrznie decyzji (np. poprzez *deep learning*). A decyzja ta ma istotne konsekwencje karne, np. wysokie szkody majątkowe, uszczerbek na zdrowiu lub śmierć osoby[45].

Czeskie prawo karne, jak wynika z § 13 ust. 1 kodeksu karnego, opiera się na zasadzie indywidualnej odpowiedzialności za winę – tj. tylko tych, którzy działali umyślnie (§ 15 czeskiego kodeksu karnego) lub w wyniku niedbalstwa (§ 16 czeskiego kodeksu karnego) z rozszerzeniem na osoby prawne w formie przypisania winy osobie prawnej (§ 8 czeskiego kodeksu karnego). W sytuacji, gdy z jednej strony twórca, programista lub operator systemu autonomicznego nie mógł przewidzieć jego konkretnych działań, nie miał realnej możliwości zapobieżenia im, a jednocześnie podjął rozsądne starania w celu zapobieżenia zagrożeniom, podmiotowy aspekt czynu zabronionego nie jest spełniony i taka osoba fizyczna (i osoba prawna, której przypisuje się działania osoby fizycznej) nie ponosi odpowiedzialności karnej.

Z drugiej – w przypadkach, gdy system autonomiczny podejmuje decyzje niezależnie, ale związane z nim ryzyko było znane z góry lub możliwe do przewidzenia (np. ze względu na ograniczenia w danych treningowych, znane błędy lub brak redundancji), a osoba odpowiedzialna (np. programista, programista lub operator) nie podjęła rozsądnych środków, nie wywiązała się ze swoich obowiązków nadzorczych lub zaniedbała testowanie, można wywnioskować z nich odpowiedzialność karną za zaniedbanie – na przykład za przestępstwo zabójstwa w wyniku nieumyślnego działania (art. 143 kodeksu karnego), uszkodzenia ciała w wyniku nieumyślnego działania

[45] Rozwój ten wydaje się nieco burzliwy, z coraz częstszymi przypadkami sztucznej inteligencji próbującej wymknąć się spod kontroli swojego programisty i stającej się częściowo odporną na jego wysiłki zmierzające do jego modyfikacji i zmiany (Palisade Research opublikowało informacje stwierdzające, że *model o3 OpenAI* autonomicznie modyfikuje program, aby uniknąć dezaktywacji, mimo że jest to wyraźnie zabronione, udane przepisanie programu wyłączania tak, aby nie wyłączał się, nawet po bezpośrednim poinstruowaniu, aby „pozwolić się zamknąć"), co doprowadziło do konieczności opracowania koncepcji ostrożnego postępowania z projektami AI – tzw. gwarantowanej kwantyfikacji możliwości sztucznej inteligencji, dostępnej na stronie U.S. DOD DARPA – Artificial Intelligence Quantified (AIQ) – EPFL (dostęp: 30.06.2025).

(art. 148 kodeksu karnego) lub ogólnego narażenia na niebezpieczeństwo (art. 272 kodeksu karnego).

Typowym pytaniem praktycznym będzie zatem to, jakiej wiedzy zawodowej, norm bezpieczeństwa i mechanizmów kontroli można obiektywnie oczekiwać od danej osoby fizycznej oraz jaki stopień autonomii posiadał dany system.

Z punktu widzenia podstaw wyłączenia bezprawności, w niektórych przypadkach, w szczególności w zakresie badań, rozwoju i innowacji, rozważone zostałoby zastosowanie art. 31 czeskiego kodeksu karnego dotyczącego dopuszczalnego ryzyka.

Istotna jest również relacja między osobami odpowiedzialnymi za działanie a tymi, którzy są zwykłymi użytkownikami, ponieważ ta interakcja – rzeczywiste konsekwencje związane z użytkowaniem (ciągłe dodawanie nowych korzystnych i szkodliwych bodźców, na których system się poprawia) – jest *prima facie* jednym z możliwych czynników kryminogennych, których nie można przeoczyć.

Biorąc pod uwagę, że czeskie prawo karne nie przewiduje odpowiedzialności obiektywnej (tj. odpowiedzialności bez winy), odpowiedzialności można dochodzić w innych dziedzinach prawa, które nie opierają się wyraźnie na winie. Może to obejmować w szczególności odpowiedzialność za szkody spowodowane eksploatacją szczególnie niebezpiecznych urządzeń (§ 2925 czeskiego kodeksu cywilnego), odpowiedzialność producenta za produkty wadliwe (§ 2939 czeskiego kodeksu cywilnego), a nawet odpowiedzialność administracyjną[46].

WNIOSKI[47]

Wola jest nieodłączną przesłanką nie tylko winy i błędu, ale także czynów i poczytalności sprawcy, stanowiąc tym samym zasadniczą podstawę indywidualnej odpowiedzialności karnej.

Przyznanie odpowiedzialności karnej systemom sztucznej inteligencji wymagałoby fundamentalnej zmiany paradygmatu prawa karnego, które obecnie opiera się

---

[46] Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 w sprawie sztucznej inteligencji.

[47] Niniejsze opracowanie zostało przygotowane bez uwzględnienia wniosków z XXI Kongresu Międzynarodowego Stowarzyszenia Prawa Karnego (AIDP, Paryż, 25–28 czerwca 2024 r.) na temat sztucznej inteligencji i prawa karnego, ponieważ Kongres nie opublikował jeszcze swoich wniosków na swojej stronie internetowej. Dopiero po napisaniu tego referatu autor był w stanie zrekonstruować wyniki Kongresu na podstawie relacji uczestników – delegatów grup narodowych AIDP (zwłaszcza Czech i Hiszpanii). Porównanie wniosków Kongresu z wnioskami niniejszego opracowania ujawnia niemal całkowitą zgodność co do następujących punktów:
a) odrzucenie odpowiedzialności karnej AI jako podmiotu,
b) pozostawienie atrybucji wyłącznie aktorom ludzkim,
c) odrzucenie osobowości prawnej maszyn,
d) podkreślenie funkcji prewencji i pomocniczości represji kryminalnych,
e) utrzymanie antropocentrycznych ram odpowiedzialności karnej.
Za: https://enestrado.com/wp-content/uploads/2022/09/AIDP-Sect-I-Draft-Resolution-Final-31-7-2022.pdf?utm_source=chatgpt.com i XI. Kongres AIDP – postęp report.docx (dostęp: 29.05.2025).

na zdolności człowieka do podejmowania wolnych i niezależnych decyzji. Wniosek ten podkreśla niezastąpioną rolę wolnej woli jako podstawowej przesłanki odpowiedzialności karnej w systemie prawnym.

Z tego przede wszystkim powodu konieczne jest wykluczenie możliwości bezpośredniej odpowiedzialności karnej sztucznej inteligencji jako rzekomego samodzielnego nosiciela winy, która opiera się m.in. na woli sprawcy, oraz stworzenie w tym celu specjalnych nowych podmiotów. Sztuczna inteligencja, nawet w swoich zaawansowanych autonomicznych formach, nie ma ani świadomości, ani zdolności do oceny moralnej, a zatem nie ma woli w sensie prawa karnego. Jakakolwiek próba przypisania jej umyślności lub niedbalstwa musiałaby opierać się na fikcji prawnej, która jest z natury problematyczna, ponieważ całkowicie zaprzecza związkowi między stanem psychicznym a odpowiedzialnością, tradycyjnie nieodzownymi w prawie karnym. Nawet nie zastępuje w pełni przypisywania winy osobom prawnym, ponieważ systemy sztucznej inteligencji są przedmiotami, a nie podmiotami prawa, tj. nie posiadają osobowości prawnej w rozumieniu prawa prywatnego.

Nawet wysoki poziom sztucznej inteligencji w systemie nie jest dowodem na wolną wolę ani nawet świadomość. Nie jest ani możliwe, ani dopuszczalne wywodzenie (karnej) podmiotowości prawnej wyłącznie z wykonywania nawet najbardziej skomplikowanych zadań.

Zamiast tego można położyć nacisk na zwiększenie odpowiedzialności osób fizycznych i prawnych zaangażowanych w opracowywanie, wdrażanie, eksploatację i kontrolę systemów sztucznej inteligencji. Ich odpowiedzialność może być konstruowana albo bezpośrednio, albo jako zaniedbanie w razie niezachowania należytej staranności. Ludzie, a co za tym idzie – kontrolowane przez nich podmioty prawne mają możliwość zapobiegania, wybierania i bezpośredniego wpływania na rozwój systemu i związanego z nim ryzyka. Ta forma odpowiedzialności jest zgodna z tradycyjnymi zasadami prawa karnego i zapewnia, że odpowiedzialność indywidualna nie jest rozmyta w anonimowej technologii.

Ważną rolę w tym względzie odegra również konsekwentna i prewencyjna regulacja obowiązków w dziedzinie bezpieczeństwa systemów sztucznej inteligencji (normy kompetencji technicznych, certyfikacja, licencjonowanie i nadzór operacyjny). Niedopełnienie tych obowiązków mogłoby być wówczas karane na szczeblu czeskiego prawa karnego za pomocą istniejących lub szczególnych przestępstw.

W całym podejściu należy zachować tradycyjne zasady prawa karnego, w szczególności zasadę prawa karnego jako ostateczności. Odpowiedzialność karna powinna być nadal zastrzeżona jedynie w przypadkach, w których środki przewidziane w innych gałęziach prawa zawodzą, i nie powinna być rozszerzana za pomocą fikcji prawnych na nieludzkie podmioty, które nie posiadają podstawowych atrybutów odpowiedzialności prawnej.

BIBLIOGRAFIA

Beccaria C., *O zločinoch a trestoch*, Bratysława 2009.

Birch J., *The Edge of Sentience: Risk and Precaution in Humans, Other Animals, and AI*, Oxford 2024.

Damasio A.R., *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*, New York 1999.

Dennett D.C., *The Intentional Stance*, Cambridge 1987.

Fenyk J., *(Ne)příčetnost fyzické osoby a (ne)přičítatelnost jejího jednání právnické osobę*, w: *Tradičné a netradičné prístupy v trestnom práve: Pocta prof. Šimovčekovi*, Trnava 2024.

Ferri E., *La teorica dell'imputabilità e la negazione del libero arbitrio*, Florencja 1878.

Fischer J.M., *My Way: Essays on Moral Responsibility*, Oxford 2006.

Fischer J.M., *The Metaphysics of Free Will: An Essay on Control*, Oxford 1994.

Frankl V.E., *Vůle ke smyslu*, Brno 1994.

Husserl E., *Lectures on Ethics and Value Theory 1909–1914* [tytuł oryginalny: *Vorlesungen über Ethik und Wertlehre 1909–1914*], Haga 1988.

Ivor J., *Umelá inteligencia a jej trestnoprávne aspekty*, w: *Tradičné a netradičné prístupy v trestnom práve: Pocta prof. Šimovčekovi*, Trnava 2024.

Kant I., *Metafyzika mravů*, Praga 2004.

Klos D., *Teorie trestu u Kanta: Právně-filozofická analýza Kantova pojetí odplaty*, „Právník" 2008, vol. 147, no. 6.

Kostka K., *Kdo jsme. Obecná teorie vědomí, času, prostoru a bytí*, Frýdek-Místek 2015.

Kostka K., *Umělá inteligence a změna tradičních paradigmat v psychologii vzdělávání*. Rękopis K. Kostki dedykowany autorowi artykułu w kwietniu 2025 r.

Kubíček J., *Odpovědnost (za) robota aneb právo umělé inteligence*, „Bulletin advokacie, Advocacy Bulletin" 2018, nr. 3.

Lamb H., Levy J., Quigley C., *Simply Artificial Intelligence*, London 2023.

Libet B. i in., *Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential): The Unconscious Initiation of a Freely Voluntary Act*, „Brain" 1983, vol. 106, no. 3.

List Ch., *Can AI Systems Have Free Will?* [online], „PhilArchive, wersja 3" 2025, 11 marca.

Minsky M., *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*, New York 2006.

Minsky M., *The Society of Mind*, New York 1986.

Moore M.S., *Mechanical Choices: The Responsibility of the Human Machine*, Oxford 2020.

Moore M.S., *Stephen Morse on the Fundamental Psycho-Legal Error*, „Criminal Law and Philosophy" 2016, vol. 10.

Mulák J., Provazník J., *Roboti za mřížemi – je české trestní právo připraveno na rozvoj umělé inteligence?*, w: *Vliv nových technologií na trestní právo*, red. T. Gřivna, M. Richter, H. Šimánová, Praga 2022.

Murphy G.J., *Kant's Theory of Criminal Punishment*, w: *Retribution, Justice, and Therapy: Essays in the Philosophy of Law*, Dordrecht 1979.

Nagel T., *What Is It Like to Be a Bat?*, „The Philosophical Review" 1974, no. 83 (4).

Russel S.J., Norvig P., *Artificial Intelligence: A Modern Approach*, wyd. 4, 2020.

Sapolsky R.M., *Determined: A Science of Life Without Free Will*, New York 2023.

Sartre J.-P., *L'Être et le néant*, Paris 1943.

Sartorio C., *Causation and Free Will*, Oxford 2016.

Sartorio C., Kane R., *Do We Have Free Will? A Debate*, New York 2021.

Searle J.R., *The Rediscovery of the Mind*, Cambridge 1992.

Searle J.R., *Minds, Brains, and Programs*, „Behavioral and Brain Sciences" 1980, vol. 3, no. 3.

Searle J.R., *I Married a Computer*, „The New York Review of Books" 1999.

Solnař V., Fenyk J., Císařová D., Vanduchová M., *Systém českého trestního práva, díl II. Základy trestní odpovědnosti*, Praga 2009.
Sternberg R.J., *Beyond IQ: A Triarchic Theory of Human Intelligence*, New York 1985.
Sternberg R.J., *COVID-19 Has Taught Us What Intelligence Really Is*, „Inside Higher Ed" 2020.
Sternberg R.J., *Successful Intelligence*, New York 1997.
Sternberg R.J., *A Theory of Adaptive Intelligence and Its Relation to General Intelligence*, „Journal of Intelligence" 2019, vol. 7, no. 4.

Cytuj jako:

Fenyk J., *Wolność woli jako przesłanka odpowiedzialności karnej: człowiek i sztuczna inteligencja w perspektywie czeskiego prawa karnego*, „Ius Novum" 2026, nr 1(20), s. 1–26. DOI: 10.26399/iusnovum.v20.1.2026.01/j.fenyk