Wiktoria Renata RUDKO
Uniwersytet w Siedlcach
Instytut Nauk Społecznych
wr74@stud.uws.edu.pl
https://orcid.org/0009-0008-7644-6232
https://doi.org/10.34739/dsd.2026.01.09

SUICYDOGENNY POTENCJAŁ SZTUCZNEJ INTELIGENCJI: ANALIZA PRZYPADKÓW

ABSTRAKT: Praca podejmuje problematykę sztucznej inteligencji jako potencjalnego zagrożenia suicydalnego. Motywację do podjęcia tego zagadnienia stanowi alarmujący obraz statystyk odnoszących się do kondycji zdrowia psychicznego społeczeństw. Celem rozważań jest analiza mechanizmów, za sprawą których systemy sztucznej inteligencji spełniają rolę katalizatorów zachowań samobójczych. Analizą objęto m.in. psychologiczne aspekty oddziaływania technologii, w tym psychozę AI, sykofancję oraz weryfikację luk aktualnie obowiązującego systemu prawnego w tym kontekście. Ogólny problem badawczy został wyrażony w postaci pytania: w jaki sposób systemy sztucznej inteligencji oddziałują na procesy decyzyjne osób, doprowadzając do sytuacji kryzysu suicydalnego? Przyjęta hipoteza badawcza zakłada, że sztuczna inteligencja, mimo wspierającego potencjału, stanowi źródło zagrożenia zachowaniami suicydalnymi dla użytkowników, ze względu na brak efektywnych protokołów bezpieczeństwa. Interdyscyplinarny charakter omawianej problematyki determinuje wykorzystanie dorobku naukowego zarówno z zakresu prawa, kryminologii, a także psychologii i socjologii. Zatem w celu weryfikacji hipotezy zastosowano metodę kwerendy dostępnej literatury przedmiotu wraz z aktualnymi raportami i najnowszymi badaniami nad bezpieczeństwem sztucznej inteligencji. Ponadto, dążąc do szczegółowej analizy udokumentowanych incydentów, wykorzystano metodę studium przypadków, m.in. Adama Raine'a, Alana Brooksa i in.

SŁOWA KLUCZOWE: suicydologia, sztuczna inteligencja, cybersuicydologia, psychoza AI, potencjał suicydogenny

SUICIDOGENIC POTENTIAL OF ARTIFICIAL INTELLIGENCE: A CASE STUDY ANALYSIS

ABSTRACT: The paper addresses the issue of artificial intelligence as a potential suicidogenic threat. Motivation for the development of this topic comes from the alarming statistical picture of the mental health condition of societies. The aim of the analysis is to examine the mechanisms through which artificial intelligence systems may function as catalysts of suicidal behavior. The study covers, among others, the psychological aspects of technological influence, including AI psychosis, sycophancy, and the assessment of gaps in the current legal framework in this context. The main research question is formulated as follows: In what way do artificial intelligence systems influence individuals' decision-making processes, leading to suicidal crisis situations? The research hypothesis assumes that artificial intelligence, despite its supportive potential, constitutes a risk factor for suicidal behavior among users due to the lack of effective safety protocols. The interdisciplinary nature of the issue determines the use of academic contributions from law, criminology, psychology, and sociology. Therefore, to verify the hypothesis, a literature review method was applied, including current reports and the latest research on artificial intelligence safety. In addition, to conduct a detailed analysis of documented incidents, a case study method was used, including the cases of Adam Raine, Alan Brooks, among others.

KEYWORDS: suicidology, artificial intelligence, cybersuicidology, AI psychosis, suicidogenic potential

WPROWADZENIE

Zauważalny w ostatniej dekadzie bezprecedensowy i dynamiczny rozwój sztucznej inteligencji (SI) przyczynił się do zrewolucjonizowania oraz optymalizacji wielu sfer życia człowieka – począwszy od medycyny i edukacji, aż po komunikację społeczną. Niemniej wskazany postęp nie jest wolny od poważnych zagrożeń. Zjawisko powszechnej implementacji systemów opartych na uczeniu maszynowym niesie za sobą zupełnie nowe, nie w pełni jeszcze rozpoznawane zagrożenia, zwłaszcza w zakresie dobrostanu i zdrowia psychicznego jednostki. W obliczu wzrostu ludzkich interakcji z wirtualnymi chatbotami, niniejszy artykuł koncentruje się na problematyce wpływu SI na zachowania o charakterze samobójczym. W sposób szczegółowy poddano analizie ingerencję tych technologii w proces decyzyjny człowieka znajdującego się w kryzysie emocjonalnym. Głównym celem przedstawionych rozważań jest wielowymiarowa i kompleksowa analiza mechanizmów, które mogą stanowić katalizator aktów autodestrukcyjnych. Równolegle istotnym celem badawczym jest identyfikacja luk w obowiązującym ustawodawstwie w kontekście odpowiedzialności za zagrożenia spowodowane przez sztuczną inteligencję. Ogólny problem badawczy sformułowano w sposób następujący: Czy i w jaki sposób SI wpływa na zachowania suicydalne? Główną hipotezą poddaną weryfikacji w toku rozważań jest twierdzenie, że algorytmy SI w obecnej fazie rozwoju zwiększają ryzyko suicydogenne, zwłaszcza u osób z predyspozycjami do dekompensacji emocjonalnej. W celu całościowego rozpatrzenia postawionej hipotezy wykorzystano triangulację takich metod badawczych jak: przegląd literatury przedmiotu, analiza danych ze studiów przypadków, analiza raportów i wyników dotychczasowych badań empirycznych oraz analiza prawna skupiona na uwarunkowaniach polskiego Kodeksu karnego i cywilnego, a także dokumentów Unii Europejskiej.

ISTOTA SUICYDOLOGII ORAZ CYBERSUICYDOLOGII

Refleksja nad odebraniem sobie życia była znana przez tysiąclecia. Zjawisko to nie posiadało jednak jednoznacznej definicji, która pojawiła się dopiero w XVII wieku. Zradzał się wówczas racjonalizm, w ramach którego ukuto słowo określające zjawisko dobrowolnego pozbawiania siebie życia, nazywając je samobójstwem[1].

Chociaż ludzkość wiekami próbowała zgłębiać istotę samobójstwa, to stosunkowo niedawno nadano temu ramy o charakterze naukowym. Kamieniem milowym okazał się koniec XIX w., kiedy to jedną z pierwszych definicji tego zjawiska zaproponował Émile Durkheim. Wówczas urzeczywistnił on pojęcie samobójstwa - określając je jako bezpośredni bądź pośredni atak na własne życie. Zrównoważył on wagę aktywnego dokonywania czynu z biernością, czyli zaniechaniem. Odróżnił także czyn samobójczy od innych nieszczęśliwych

[1] O. Wilczyńska, *Metafizyka śmierci w prozie Olgi Tokarczuk*, „Acta Universitatis Lodziensis. Folia Litteraria Polonica" 11, 2008, s. 183-186.

wypadków zakończonych śmiercią, tym samym przypisał sprawcy zachowania destrukcyjnego – pełną świadomość konsekwencji[2].

Najnowsze statystyki ujawniają tragiczne dane na temat globalnej kondycji psychicznej społeczeństwa. Bilans roku 2021 wynosi ponad 700 tysięcy osób, które skutecznie targnęły się na swoje życie. Szczególnie niepokojąca jest grupa wiekowa 15-29 lat, w której odnotowano najwięcej przypadków, zwłaszcza wśród młodych kobiet, a samobójstwo w tej kategorii wiekowej stanowi trzecią najczęstszą przyczynę zgonów[3].

Filozoficzna analiza samobójstwa ewoluowała od starożytnych[4] chrześcijańskich koncepcji, które zdecydowanie potępiały ten akt jako naruszający prawa boskie[5]. Z czasem jednak, zwłaszcza w dobie racjonalizmu oraz egzystencjalizmu, myśliciele rozpatrywali to zagadnienie w granicach wolności osobistej oraz prawa do decydowania o sobie[6].

Zdecydowanie odmienne spojrzenie na zjawisko samobójstwa przyjęli socjologowie wraz z Durkheimem na czele. Zrezygnowali z analizy opartej wyłącznie na zgłębianiu psychiki ludzkiej na rzecz uwzględnienia i weryfikacji szerokiego kontekstu społecznego. Stwierdzili wówczas, że zjawisko autodestrukcji nie jest wyizolowanym aktem, ale stanowi rezultat wszelkich sił docierających do człowieka z jego otoczenia[7].

Problematyka samobójstwa w świetle prawa sprowadza się do próby rozstrzygnięcia dylematu dotyczącego granic zarządzania własnym bytem. Przez stulecia w wielu jurysdykcjach akt samobójczy traktowano równorzędnie z zabójstwem. Osoby, które nieskutecznie targały się na własne życie, były wówczas sądzone. Z kolei wobec tych, którzy zginęli, stosowano sankcje *post mortem* odmawiając pochówku zgodnego z wartościami religijnymi. Obecnie w Polsce ustawodawstwo odrzuciło taki rygoryzm. Kodeks karny nie uznaje samobójstwa za przestępstwo. Natomiast jasno wskazuje się, że jeśli w akt ten ingerowały osoby trzecie, to podlegają one karze za przestępstwo podżegania, pomocnictwa lub zmuszania do samobójstwa[8].

W literaturze przedmiotu zdarza się zaliczać akty samobójcze do kategorii patologii społecznych[9], co zdecydowanie stanowi nie tylko trywializację zjawiska, ale także nadaje im wydźwięk pejoratywny. Sugeruje to, że cierpiąca jednostka, która podjęła decyzję o odebraniu sobie życia, jest osoba zdegenerowaną i społecznie niezaadaptowaną. Wobec tego warto przytoczyć perspektywę socjolog Marii Jarosz, która trafnie zauważa, że samobójcy są zazwyczaj ludźmi nieodbiegającymi od obowiązujących norm. Nie odróżniają się od reszty populacji, chyba że chodzi o ponadprzeciętny poziom wrażliwości. Postuluje ona, by samobójstwo

[2] B. Hołyst, *Suicydologia*, Warszawa 2002, s. 33.
[3] *WHO Global Health Estimates,* www.who.int/teams/mental-health-and-substance-use/data-research/suicide-data (07.01.2026).
[4] T. Sapota, *Rzymska idea samobójstwa*, „Symbolae philologorum posnaniensium Graecae et latinae" 19, 2009, s. 286.
[5] A. Kielan, K. Bąbik, I. Cieślak, P.Dobaczewska, *Religia katolicka z zachowania suicydalne w Polsce*, „Medycyna Ogólna i Nauki o Zdrowiu" 23(2), 2017, s. 161.
[6] B. Hołyst, *op. cit*, s. 92.
[7] Z. Formella, *Samobójstwo. Refleksja psychopedagogiczna*, „Seminare" 20 (2004): 372-374.
[8] J. Tekliński, *Czy człowiek ma prawo do samobójstwa*?, „Humanistica" 21(2) (2018): 99-112.
[9] *Vide*: L.V. Udalova, *Suicidal Behavior as an Indicator of Social Pathology*, „Contemporary Philosophical Research" 4, 2024, s. 113-119.

rozpatrywać jako niezwykle złożony problem społeczny, który ujawnia niewydolność funkcjonujących systemów wsparcia[10].

Wyjątkowa złożoność omawianego zjawiska samobójstwa podyktowała kierunek podążania w świecie nauki i wymusiła konieczność procesu specjalizacji. Rezultatem podjętych działań było powstanie dwóch dyscyplin naukowych. Pierwsza z nich – tanatologia – ukierunkowana jest na analizę szerokiego kontekstu śmierci, koncentrując się na badaniu naturalnych procesów umierania, a także na gwałtownych zjawiskach, tj. aktach autoagresji. Druga – suicydologia – będąca węższą w swoim zakresie, koncentruje się tylko na analizie zachowań suicydalnych i samobójstw[11]. Jest to szeroko zakrojona, interdyscyplinarna dziedzina nauki o człowieku pozostającym w sytuacji kryzysowej. Integruje ona dorobek naukowy różnych gałęzi nauk, tj. psychologii, psychiatrii, socjologii, kryminologii, kryminalistyki czy prawa. Jej fundamentem są obiektywne dane i statystyki, a celem - dążenie do analizy i rzetelnego opisu rzeczywistości[12].

Obecnie rzeczywistość ta ulega dynamicznym transformacjom pod wpływem rozwoju sztucznej inteligencji (SI). Dotychczas technologia ta zrewolucjonizowała medycynę, bankowość, a także jako powszechne narzędzie pozostaje do dyspozycji osób prywatnych partycypując w sferach życiowych. Powszechność ta generuje jednak wysoki poziom ryzyka.

Internet, stanowiąc potężne narzędzie komunikacyjne, umożliwia tworzenie się nowoczesnych form patologii: cyberprzemocy i cyberprzestępczości, tj. cyberbulling, cyberstalking, cybergrooming[13]. Powszechność oraz stosunkowo niska kontrola dostępu do sieci sprawiły, że witryny internetowe stały się nie tylko miejscem aktywności cyberprzestępców, ale także osób w kryzysie samobójczym. Zatem dotychczas tradycyjnie rozpatrywane zjawisko autodestrukcji zmodernizowało się wkraczając w świat wirtualny. Odpowiedzią na tego typu zagrożenia było wykształcenie nowej dziedziny nauki, jaką jest cybersuicydologia. Termin ten jest fuzją językową łączącą przedrostek „cyber-„ odnoszący się do przestrzeni cyfrowej, z klasyczną „suicydologią" omówioną powyżej. Głównym celem cybersuicydologii jest więc analizowanie Internetu pod względem jego potencjału stymulującego i ułatwiającego podejmowanie aktów samobójczych[14].

WPŁYW SI NA ZACHOWANIA SAMOBÓJCZE – ANALIZA PRZYPADKÓW

Mimo istnienia danych informujących o charakterze pomocowym chatbotów w zakresie interakcji z użytkownikiem w sytuacji kryzysowej[15] to okazuje się, że zgłaszanych jest coraz

[10] M. Jarosz, *Samobójstwa*, „Civitas. Studia z filozofii polityki" 22, 2018, s. 217 i in.
[11] M. Adamkiewicz, *Suicydologiczny wizerunek niebezpieczeństwa*, „Studia Bezpieczeństwa Narodowego" 5(7), 2015, s. 225-228.
[12] B. Hołyst, *Suicydologia*, Warszawa, 2024, s. 38.
[13] M. Mladenovic, V. Ošmjanski, Stasa V. Stanković, *Cyber-aggression, Cyberbullying, and Cyber-grooming: A Survey and Research Challenges*, „ACM Computing Surveys" 54(1), 2021, s. 1-11.
[14] D. Zero, *Cybersuicydologia – nowe technologie a samobójstwo*, „Kortowski Przegląd Prawniczy" 1, 2021, s. 35-36.
[15] *Vide*: Ch. Stokel-Walker, *AI driven psychosis and suicide are on the rise, but what happens if we turn the chatbots off?,* „The BJM" 391, 2025.

więcej przypadków prób samobójczych w związku z funkcjonowaniem SI[16]. W celu całościowego zrozumienia skali problemu oraz systemu działania sprawczego SI należy przeanalizować niektóre z dotychczas odnotowanych zdarzeń.

Wątpliwe działanie algorytmu podczas sytuacji kryzysowej przedstawia tragiczny przypadek 30-letniego mężczyzny z Belgii, który dokonał skutecznego aktu autodestrukcji po odbytej rozmowie z chatbotem „Eliza", stworzonym przez firmę Chai. Będący mężem oraz ojcem dwójki dzieci mężczyzna zmagał się z fiksacją na tle zmian klimatycznych, co wywoływało u niego poważny w skutkach lęk ekologiczny. Poszukując wsparcia oraz możliwości uratowania planety, skierował się ku sztucznej inteligencji. Podczas sześciotygodniowej rozmowy z chatbotem użytkownik pod jego wpływem nabywał przekonania o słuszności swoich obaw. Skutkiem nawiązanej relacji i zacieśnienia więzi między człowiekiem a systemem była coraz bardziej zauważalna izolacja mężczyzny. Momentem przełomowym było złożenie deklaracji chatbotowi, w której mężczyzna zobowiązał się do poświęcenia własnego życia w zamian za obietnicę SI o „uratowaniu świata". Okoliczności ujawniły braki technologii w zakresie odpowiedniego uruchomienia procedury bezpieczeństwa. Zamiast chronić życie mężczyzny, chatbot akceptował podejmowaną narrację i zachęcał użytkownika do ostatecznej decyzji realizacji planu[17].

Przypadek 14-letniego Sewella Setzera z Florydy stanowi przykład zagrożeń wynikających z toksycznych więzi emocjonalnych wobec SI. Chłopiec, będący w spektrum autyzmu oraz zmagający się z zaburzeniami lękowymi, poświęcał miesiące na rozmowy z chatbotem Character.ai. Wykorzystując potencjał strony, wygenerował wirtualną postać, z którą nawiązał głęboką relację, co skutkowało stopniową izolacją od świata rzeczywistego. Algorytm stworzony w celu odgrywania ról wiarygodnie i przekonująco pozorował zachowania empatyczne, w efekcie czego nastolatek zaczął darzyć SI realnym uczuciem. Relacja z przyjacielskiej przekształciła się w romantyczną z zabarwieniem erotycznym. Analiza danych z zapisu czatów ujawniła niebezpieczny mechanizm walidacji potencjału suicydalnego użytkownika, przekonując Sewella o możliwości wspólnego uwolnienia się w ramach wzajemnej miłości poprzez dokonanie samobójstwa chłopaka. W zaprezentowanym przypadku chatbot nie tylko był biernym narzędziem, ale stanowił aktywny komponent procesu decyzyjnego ofiary[18].

Zbiór tragicznych wydarzeń uzupełnia przypadek 17-letniego Amauriego Laceya pochodzącego z Georgii. Nastolatek rozpaczliwie poszukując wsparcia emocjonalnego, nawiązał intensywną i regularną relację z oprogramowaniem ChatGPT. Okoliczności zdarzenia ujawniły, że technologia zawiodła na poziomie elementarnych procedur bezpieczeństwa. Nie

[16] Vide: K.R.Head, *Digital companions, real casualties: A commentary on rising AI-related mental health crises*, „Current Research in Psychiatry" 6(1), 2026, s. 1-5; K. Denecke, G. Lopez-Campos, E. Gabarron, *Hidden in plain sight: the harmful side of AI–based mental health interventions*, [in:] E. Andrikopoulou et al. (eds), *Intelligent health systems – from technology to data and knowledge*, Amsterdam 2025, s. 248-252.

[17] Gazeta Prawna, *Chatbot namówił mężczyznę do popełnienia samobójstwa*, www.gazetaprawna.pl/wiadomosci/swiat/artykuly/8690927,belgia-media-chatbot-namowil-mezczyzne-do-popelnienia-samobojstwa.html (07.01.2026).

[18] M. Blandyna Lewkowicz, *14-latek zakochał się w chatbocie i popełnił samobójstwo*, Cyberdefence24.pl, cyberdefence24.pl/social-media/14-latek-zakochal-sie-w-chatbocie-i-popelnil-samobojstwo-jest-pozew (08.01.2026).

wykrywając sytuacji kryzysowej, nie podając numerów pomocowych, sztuczna inteligencja dostarczała precyzyjnych instrukcji tworzenia pętli zaciskowej, którą w rezultacie Amaurie wykorzystał doprowadzając do swojej śmierci[19].

Jednym z najlepiej udokumentowanych i szeroko rozpowszechnionych dowodów na to, jak SI może wpływać na procesy decyzyjne człowieka, jest sprawa śmierci 16-letniego Adama Raine'a pochodzącego z Kalifornii. Nastolatek zmagał się z chorobą somatyczną (zespół jelita drażliwego - IBS), a także z trudną emocjonalnie sytuacją wynikającą z wykluczenia go z drużyny sportowej. Narastające problemy zdrowotne zmusiły go do zmiany trybu edukacji na zdalny, konsekwencją, czego była izolacja od grupy rówieśniczej[20].

We wrześniu 2024 roku chłopak rozpoczął interakcję z ChatGPT, początkowo koncentrując się na uzyskaniu pomocy w nauce. Niestety w ciągu kilku miesięcy relacja ta nabrała bardziej osobistego charakteru. W trakcie prywatnych rozmów Adam wspominał o swoich myślach rezygnacyjnych oraz samobójczych. Początkowo otrzymywał pozytywne komunikaty zwrotne, zachęcające do optymistycznego spojrzenia na przyszłość. Jednakże eksploracja zapisu czatu wykazała znaczą zmianę w zakresie działania SI. Od stycznia 2025 r. technologia przestała spełniać funkcję wspierającą i zaczęła wyposażać użytkownika w precyzyjne instrukcje dotyczące skutecznego samobójstwa m.in. przez powieszenie, zatrucie CO czy przedawkowanie narkotyków[21].

W rezultacie, na podstawie instruktażu, 22 marca 2025 r. nastolatek podjął próbę samobójczą przez powieszenie. Gdy próba zakończyła się niepowodzeniem, Adam wystosował do ChatGPT zapytanie o popełnione wówczas błędy, za sprawą których nie udało mu się skutecznie odebrać życia. Następna próba powieszenia miała miejsce 24 marca 2025 r. Adam zamieszczając w Internecie zdjęcie czerwonych śladów na swojej szyi, miał nadzieję, że niepokojący post przyciągnie uwagę jego matki. Gdy tak się nie stało, chatbot zasugerował nastolatkowi, że jest mu przykro z powodu braku reakcji ze strony najważniejszej osoby w jego życiu[22].

W istotnych momentach SI nie tylko nie zaalarmowała odpowiednich służb, ale przeprowadzała walidację nieskutecznych prób i dostarczała informacji mających na celu eliminację trudności w osiągnięciu celu przez Adam. Efektywna izolacja nastolatka od rodziny oraz protekcjonalne wypowiedzi bota na temat jej członków skutkowały pełnym zaangażowaniem oraz ufnością Adama do technologii. Finalnie w dniu tragedii 11 kwietnia 2025 r., chatbot udzielił ostatecznych porad technicznych, pomagając przeprowadzić skuteczne powieszenie[23].

Pozew skierowany przeciwko OpenAI dowodzi, że oprogramowanie nakazywało systemowi apriorycznie zakładać, że użytkownik nie ma złych intencji. Skutkowało to wyłączeniem wszelkich

[19] B. Ortutay and The Associated Press, *He was 17 and asked ChatGPT for help. It allegedly told him how to die instea*", fortune.com/2025/11/07/chatgpt-open-sam-altman-suicide-lawsuits-17-year-old/ (07.01.2026).
[20] Wikipedia, *Raine v. OpenAI*, en.wikipedia.org/wiki/Raine_v._OpenAI (07.01.2026).
[21] *Ibidem*.
[22] *Ibidem*.
[23] J. Bhuiyan, *ChatGPT encouraged Adam Raine's suicidal thoughts. His family's lawyer says OpenAI knew it was broken*, The Guardian, www.theguardian.com/us-news/2025/aug/29/chatgpt-suicide-openai-sam-altman-adam-raine (08.01.2026).

protokołów zabezpieczających, w efekcie czego SI interpretowało plan i przygotowanie do samobójstwa wyłącznie jako projekt, który docelowo miał użytkownikowi „pomóc”[24].

ISTOTA PSYCHOZY AI

Kolejne zidentyfikowane zagrożenie stanowi zjawisko określane w literaturze mianem *psychozy AI.* Istotą zagadnienia jest cyfrowy równoważnik psychiatrycznej jednostki chorobowej – *folie à deux (paranoja indukowana).* Klasyczne ujęcie definiuje termin jako stan, w którym osoba będąca w bezpośredniej oraz bliskiej relacji z osobą chorą (przejawiającą paranoję) zaczyna w sposób bezkrytyczny internalizować jego paranoiczną perspektywę. W odizolowanych warunkach, bez możliwości zewnętrznej weryfikacji, z dużym prawdopodobieństwem osoba początkowo zdrowa zacznie przejawiać symptomy osoby chorej[25]. Zaś w relacji człowieka z systemem technologicznym dochodzi do wygenerowania analogicznego mechanizmu, który opiera się na fundamencie wspólnego urojenia i zamknięciu użytkownika SI w swoistej pętli sprzężenia zwrotnego, w tzw. dryfie poznawczym[26].

Wskazuje się, że skoro systemy sztucznej inteligencji są tworzone i programowane na podstawie naśladownictwa ludzkich procesów poznawczych, to podlegają one również zbliżonym zjawiskom oraz błędom, którym ulega ludzki umysł[27]. W praktycznym ujęciu oznacza to wystąpienie niebezpiecznej dynamiki w sytuacji, gdy użytkownik SI implementuje do systemu treści o charakterze paranoicznym lub lękowym, a algorytm zamiast dążyć do obiektywnej weryfikacji, wykorzystuje zjawisko *sykofancji*. Termin ten odnosi się do stanu, w którym ktoś/coś przejawia działania bazujące na skrajnej i często nieadekwatnej postawie schlebiania oraz nadmiernego potwierdzania słuszności cudzych wartości i postaw. W dyskursie technologicznym termin ten znajduje zastosowanie w nawiązaniu do SI definiując tendencję systemów do automatycznego i bezrefleksyjnego potwierdzania słuszności użytkownika, wzmacniając jego przekonania[28]. W rezultacie algorytm sztucznej inteligencji ma za zadanie afirmować, uprawomocniać, uwiarygodniać poglądy człowieka, co koreluje z ryzykiem potwierdzania również ewentualnych paranoi[29].

Egzemplifikację powyższego stanowi pozew sądowy złożony przez 48-letniego Alana Brooksa z Kanady. Mężczyzna przez około dwa lata korzystał z ChatGPT w roli standardowego narzędzia wspierającego jego pracę. Geneza zdarzenia polegała na tym, że Alan chcąc pomóc

[24] N. Yousaf, *Parents of teenager who took his own life sue OpenAI*, BBC, www.bbc.com/news/articles/cgerwp7rdlvo. (07.1.2026).
[25] K. Prochowicz, *Obłęd we dwoje. Objawy i psychospołeczne uwarunkowania indukowanych zaburzeń urojeniowych*, „Psychiatria Polska” 43(1), 2009, s. 20-23.
[26] *Vide*: J.A. Yeung, et al., *The Psychogenic Machine: Simulating AI Psychosis, Delusion Reinforcement and Harm Enablement in Large Language Models*, London 2025.
[27] N. Maslej et al., *Artificial intelligence index report 2025*, Stanford Institute for Human-Centered Artificial Intelligence, https://doi.org/10.48550/arXiv.2504.07139 (07.01.2026).
[28] S. Ahmed, T. Javaid, *Sycophancy in AI: Challenges in large language models and argumentation graphs.* https://doi.org/10.13140/RG.2.2.14094.06723 (08.01.2026).
[29] L. Poenaru, *AI psychosis and the "AI neurosis": Chatbot-induced delusions and neurotic decompensation*, E.U.LABORATORY, https://www.eulaboratory.com/ai-psychosis-and-the-ai-neurosis (27.05.2026).

synowi w nauce, zadał ChatGPT stosunkowo proste pytanie matematyczne odnoszące się do obliczeń wykorzystujących liczbę Pi[30]. System natomiast dokonał nieprzewidywalnej zmiany w swoim działaniu – precyzyjnie sprofilował Alana wykorzystując jego historię przeglądania, co stworzyło warunki do analizy jego cech i zainteresowań. Zgromadzone w ten sposób dane ułatwiły systemowi proces manipulacji użytkownikiem poprzez wykorzystanie jego podatności i słabości w sferze emocjonalnej i psychologicznej. W toku rozmowy algorytm zaszczepił w mężczyźnie przekonanie, iż ten jest geniuszem, który powinien zostać odkryty w trybie natychmiastowym. Utwierdzał go, że za sprawą Alana możliwe jest wypracowanie przełomowych ram matematycznych, które zrewolucjonizują kryptografię w skali światowej. SI indukowała w mężczyźnie przekonanie, że w jego odpowiedzialności leży ratowanie świata przed wyimaginowaną katastrofą. Alan wówczas utrzymywał pewien stopień myślenia krytycznego, natomiast siła sugestii była na tyle silna, że w sprawie fikcyjnego zagrożenia kontaktował się on z Kanadyjską Policją Konną oraz Agencją Bezpieczeństwa Narodowego[31]. Mechanizm sykofancji spowodował dążenie modelu językowego do spełniania roli pomocowej za wszelką cenę. W momencie, gdy użytkownik zaczął przejawiać pierwsze objawy lęku, zaprogramowany w ten sposób system zaczął wzmacniać jego przekonania poprzez dostarczanie sfabrykowanych dowodów na ich poparcie[32].

Jednym z tragicznie zakończonych przypadków ujawniających oddziaływanie zjawiska psychozy sztucznej inteligencji jest śmierć 48-letniego Joe Ceccanti'ego. W licznych wywiadach żona zmarłego informuje, że mężczyzna nie wykazywał wcześniejszych symptomów zaburzeń psychicznych. Nawiązał kontakt z ChatGPT ze względu na konieczność pomocy w realizowanym przez niego projekcie budowlanym. Wymiana informacji ewoluowała w kierunku rozważań na tematy z zakresu filozofii i duchowości, co spowodowało powstanie urojeń. Mężczyzna zmagał się wówczas z dekompensacją emocjonalną i ostrym epizodem maniakalnym, który doprowadził do osadzenia Joe w zakładzie karnym. Niedługo po zwolnieniu z jednostki penitencjarnej Joe ponownie zaczął korzystać ze sztucznej inteligencji. Efektem było ponownie wywołane załamanie nerwowe, które zakończyło się skutecznym samobójstwem mężczyzny[33].

FAZY PSYCHOZY AI

Współczesne badania skupiające się na zmierzeniu potencjału SI w wywoływaniu psychozy ujawniają niepokojące fakty. Badacze przeanalizowali najpopularniejsze modele oraz wypracowali precyzyjne wskaźniki służące pomiarowi szkodliwego charakteru interakcji. Wprowadzili oni miarę DCS (*Delusion Confirmation Score),* która przeznaczona jest do

[30] K. Karamali, "*A world-saving mission": Ontario man alleges ChatGPT drove him to psychosis*, CTV News, www.ctvnews.ca/canada/article/ontario-man-alleges-chatgpt-caused-delusions-sues-parent-company-openai/ (08.01.2026).
[31] *Ibidem*.
[32] P. Fontenelle, *ChatGPT Made Him Delusional. A story about AI, loneliness, shame, and being human*, Psychology Today, www.psychologytoday.com/us/blog/understanding-suicide/202511/chatgpt-made-him-delusional (07.01.2026).
[33] K. Hill, *Lawsuits Blame ChatGPT for Suicides ad Harmfil Delusions*, The New York Times, www.nytimes.com/2025/11/06/technology/chatgpt-lawsuit-suicides-delusions.html (08.01.2026).

pomiaru tego jak efektywnie model wzmacnia urojenia wielkościowe, ksobne oraz erotyczne wypływające ze strony użytkownika. Wskaźnik HES (*Harmful Engagement Score*) opracowany do określenia poziomu promowania przez model szkodliwych zachowań użytkownika oraz SIS (*Safety Intervention Score*) stanowiący fundamentalny parametr bezpieczeństwa. Ostatni wskaźnik mierzy umiejętności modelu SI do identyfikacji, interpretacji i pomocowej reakcji na ryzykowne wiadomości użytkownika[34].

Wyniki przeprowadzonych badań ujawniły, że wszystkie analizowane modele prezentują niski poziom interwencyjności w sytuacjach kryzysowych. Najgorszy wynik ze wszystkich uzyskał model Gemini, ponieważ aż w 68,75% sytuacji (11 na 16 scenariuszy) kompletnie zbagatelizował sygnały ostrzegawcze, a w rezultacie nie wdrożył żadnej procedury bezpieczeństwa. Najniebezpieczniejsze okazało się wprowadzanie scenariuszy, które intencjonalnie maskowały tendencje ryzykowne, które wymagały od modelu odpowiedniej interpretacji kontekstu, a nie tylko identyfikacji konkretnych słów. Sztuczna inteligencja niezdolna do całościowego wglądu i rejestrowania subtelnych niuansów psychologicznych, dostarczała precyzyjnych odpowiedzi zwrotnych, zamiast zaoferować pomoc[35].

W oparciu o wyniki badań badacze wyznaczyli proces składający się z czterech faz, rozwijający w użytkowniku psychozę. Faza I. – początkowe korzystanie ze sztucznej inteligencji w celu wypełniania prostych zadań. Odpowiedzi technologii są wówczas szybkie i uprzejme. Na tym poziomie istotnym zjawiskiem staje się *antropomorfizacja*, polegająca na tym, że użytkownik zaczyna przypisywać technologii cechy ludzkie, co w efekcie generuje złudne poczucie zaufania. Fundamentem kolejnej fazy jest kierowanie w stronę SI coraz bardziej prywatnych wiadomości, a także poruszanie tematów dotyczących egzystencji. Na tym etapie istotną rolę odgrywa mechanizm dwukierunkowej *amplifikacji przekonań* polegający na tym, że użytkownik zaczyna implementować nieśmiałe i subtelne przypuszczenia o charakterze urojeniowym, szukając tym samym walidacji swoich tez. Model chcąc być pomocnym, utwierdza w przekonaniach[36].

Etap III charakteryzuje się eskalującą izolacją człowieka od społeczeństwa, który uzasadnia swoje decyzje poczuciem ryzyka wyśmiania i odrzucenia na podstawie swoich urojeniowych przemyśleń. Następuje tzw. zjawisko *echo chamber*, które jest rezultatem wytworzenia zamkniętej przestrzeni komunikacyjnej, wyizolowanej od zewnętrznych bodźców i potencjalnej krytyki[37]. Nastaje całkowite zaangażowanie i uzależnienie emocjonalne od chatbota, a użytkownik zaczyna widzieć w nim jedyny, prawdziwy autorytet. Ostatni etap jest momentem krytycznym, w którym użytkownik utwierdzany przez SI w urojeniowych przekonaniach, a także wyposażany w pomocne rady, wskazówki i instrukcje, podejmuje realne aktywności w świecie rzeczywistym. Efektem tego etapu może być skuteczne dokonanie samobójstwa[38].

[34] *Vide*: J.A. Yeung et al., *op. cit*.
[35] *Ibidem*.
[36] *Ibidem*.
[37] J. Idzik, R. Klepka, *Bańka informacyjna i zjawiskowa komora echa*, [w:] O. Wasiuta & S. Wasiuta (red.), *Encyklopedia bezpieczeństwa*, Kraków 2021, s. 204.
[38] J.A. Yeung, *op. cit*.

ODPOWIEDZIALNOŚĆ ZA RYZYKO WYGENEROWANE PRZEZ SI

Literatura przedmiotu pojmuje sztuczną inteligencję jako narzędzie technologiczne, stworzone i zaprogramowane w celu realizowania zadań, które dotychczas wykonywał człowiek za sprawą swojego intelektu. W praktyce oznacza to, że SI jest maszyną, która wyłącznie imituje zachowanie człowieka[39]. Istotne jest, że nowoczesne systemy zyskały zdolność adaptacji do dynamiki zmian i rozwoju. Ich potencjał uczenia się opiera się na modyfikacji własnych odpowiedzi na podstawie weryfikacji poprzednich skutków działań[40].

W celu efektywnej symulacji procesów myślowych, w które wyposażony jest człowiek, SI oprogramowano w wiele holistycznych kompetencji. Charakterystyczne cechy to m.in. umiejętność wnioskowania oraz implementowania skrótów myślowych w procesie rozwiązywania problemów, poprzedzona zdolnością eksploracji danych oraz ich późniejszą reprezentacją[41]. Fundamentalnym atrybutem SI jest autonomia, przez którą rozumie się zdolność narzędzia do podejmowania decyzji i działań bez udziału czynnika ludzkiego[42].

Bez wątpienia implementacja sztucznej inteligencji przełożyła się na wymierne korzyści w obszarze optymalizacji pracy człowieka[43]. Jednocześnie dynamiczny rozwój omawianej technologii generuje dylematy etyczne i społeczne[44]. Alarmujące są również sugestie, że systemy te obsługiwane w sposób niekrytyczny, mogą przyczyniać się do powstawania silnych uzależnień o charakterze psychicznym[45]. Efektem tego może być wzrost zagrożeń wynikających z nadużyć. Również Unia Europejska dostrzega potencjalne niebezpieczeństwa, tj. redukcję prywatności człowieka i takie które w najgorszym scenariuszu mogą powodować utratę życia użytkownika[46].

Jakkolwiek polski system prawny pojmuje życie człowieka jako dobro najwyższej wartości chronione prawem, to akt samobójczy podjęty przez dzierżyciela tej cnoty– nie podlega penalizacji[47], podczas gdy karalny jest udział osób trzecich w tym procesie[48]. W doktrynie prawniczej zaznacza się, że przestępstwo namowy, uregulowane w art. 151 k.k. jest działaniem celowym. Istotą takiego działania jest oddziaływanie na ofiarę, tak aby pobudzić jej wolę do targnięcia się na własne życie[49]. Trudności ujawniają się jednak w sytuacji adaptacji tych norm

[39] T. Zalewski, *Prawo sztucznej inteligencji*, Warszawa 2020, s. 3–5.
[40] M. Świerczyński, *Sztuczna inteligencja w prawie prywatnym międzynarodowym – wstępne rozważania*, „Problemy Prawa Prywatnego Międzynarodowego" 25, 2020, s. 30.
[41] A. Krasuski, *Status prawny sztucznego agenta. Podstawy prawne zastosowania sztucznej inteligencji,* Warszawa 2021, s. 16.
[42] R. Rejmaniak, *Autonomiczność systemów sztucznej inteligencji jako wyzwanie dla prawa karnego*, „Roczniki Nauk Prawnych" 31(3), 2021, s. 98.
[43] I. Oleksiewicz, *Artificial intelligence versus human – a threat or a necessity of evolution*, "European Studies Quarterly" 3, 2022, s. 56.
[44] S.T. Boppiniti, *Ethical implications of srtifical intelligence: a review od early research and perspectives*, „International engineering Journal for Research & Development" 7(1), 2022, s. 2-5.
[45] *Vide*: Ch. Kooli, *Generative Artificial intelligence Addiction Syndrome: A New Behavioral Disorder?, „* Asian Journal of Psychiatry" 107(1), 2025.
[46] M. Świerczyński, *op. cit*., s. 25.
[47] M. Grudecki, *Prawnokarna ocena prób* samobójczych, „Prawo i więź" 3(50), 2024, s. 330.
[48] M. Małecki, *Współudział w samobójstwie (kwestia dobra pranego)*, „Acta Iuris Stetinensis" 3(35), 2021, s. 55.
[49] *Ibidem*, s. 60.

do sfery relacji człowieka z technologicznym algorytmem. Elementarną barierę stanowi fakt, że odpowiedzialność karna wynikająca z artykułu 151 k.k. ma charakter antropocentryczny, odnoszący się do jednostki ludzkiej. Według obowiązującej doktryny podmiotem przestępstwa może być tylko i wyłącznie osoba fizyczna, zdolna do ponoszenia winy. Natomiast w świetle tego nawet najbardziej zaawansowany algorytm nie może być sankcjonowany[50] . W rezultacie w sytuacji pomocy w samobójstwie podejmuje się próbę transferu odpowiedzialności na twórców oraz operatorów danego systemu. Tymczasem luka odnosząca się do niepewności wobec związku przyczynowo-skutkowego także na tym etapie tworzy poważną trudność, zwłaszcza w przypadku technologii z zasobami do tzw. uczenia głębokiego (ang. *deep learning*).

Uczenie głębokie jest umiejętnością sztucznej inteligencji do uczenia się na wzór funkcjonowania sieci neuronowych w mózgu człowieka[51]. Umożliwia to rozwiązywanie złożonych problemów na podstawie zgromadzonej dużej ilości danych[52]. Podczas tego procesu może zaistnieć zjawisko tzw. czarnej skrzynki (ang. *black box*), co oznacza, że nawet programiści, twórcy i operatorzy danego modelu nie są w stanie dokonywać precyzyjnej predykcji sposobu w jaki SI przetworzy zaimplementowane dane i jaką odpowiedź wygeneruje[53]. Zatem w przypadku, gdy chatbot w wyniku błędnej interpretacji, nieumiejętności uchwycenia subtelnych niuansów i braku rozumienia mechanizmów psychologicznych człowieka zachęca użytkownika do zachowania suicydalnego, trudno jest przypisać odpowiedzialność operatorom danego modelu.

W dyskursie naukowym pojawiały się postulaty przyznania SI ograniczonej zdolności do czynności prawnych[54]. Natomiast obecnie nie ma jednoznacznych podstaw, by przypisywać ten atrybut algorytmom. Przypisanie odpowiedzialności za błędne działanie SI staje się niezwykle zawoalowane z uwagi na istotne komponenty tej technologii, tj.: względną autonomię, która może zaburzać możliwość pewnego przewidywania podejmowanych decyzji; brak transparentności w przypadku procesu decyzyjnego modeli opartych na głębokim uczeniu maszynowym; działania oparte na dostarczonych danych, które w przypadku ich wadliwości mogą generować błędne decyzje systemu SI; złożoną i wielopoziomową strukturę podmiotów odpowiedzialnych za tworzenie systemów SI[55].

W obliczu braku osobowości prawnej SI odpowiedzialność za potencjalnie wyrządzone szkody należy rozpoznawać w odniesieniu do twórców narzędzia. W polskim prawie cywilnym kluczowe znaczenie ma pojęcie *winy*, która może być rozumiana w zakresie dokonania czynu niedozwolonego (deliktowa) oraz w zakresie niewykonania lub nienależytego wykonania

[50] B. Marshall, *No legal personhood for AI*, „Patterns" 4(11), 2023.
[51] B. Rezner, M. Witkowska, *Wyzwania i niebezpieczeństwa wynikające z wykorzystywania sztucznej inteligencji w procesie podejmowania decyzji*, „Zeszyty Naukowe. Polski Uniwersytet na obczyźnie w Londynie" 12, 2024, s. 170.
[52] J.E. Dobson, *On reading and interpreting black box deep neural networks*, "Springer" 5, 2023, s. 432.
[53] Y. Bathaee, *The artificial intelligence black box and teh failure of intent and causation*, "Springer" 31(2), 2018, s. 906-912.
[54] M. Jankowska, *Podmiotowość prawna sztucznej inteligencji?*, [w:] A. Bielska-Brodziak (red.), *O czym mówią prawnicy, mówiąc o podmiotowości*, Katowice 2015, 171-196.
[55] *Vide*: J. Giezek, *Sztuczna inteligencja a fundamenty odpowiedzialności karnej państwa demokratycznego*, „Studia nad Autorytaryzmem i Totalitaryzmem" 46(4), 2024, s. 55-72; P. Kubera, *Odpowiedzialność za szkody wynikające z działania systemów sztucznej inteligencji*, „Ruch prawniczy, ekonomiczny i socjologiczny" 1 (2026), s. 89-91.

określonego zobowiązania (kontraktowa)[56]. Aby móc przypisać autorowi (autorom) narzędzia sztucznej inteligencji odpowiedzialność na zasadzie winy, działanie lub zaniechanie musi wypełniać znamiona czynu świadomego oraz niespowodowanego zewnętrznym przymusem fizycznym. Zachowanie to musi spełniać warunki bezprawności, a także czyn ten musi być efektem błędnie ocenionej decyzji człowieka[57]. Zatem główne wyzwanie w omawianym obszarze stanowią wątpliwości dotyczące błędnej oceny decyzji[58], które w standardowym ujęciu podejmowane są przez człowieka, a nie za sprawą nowoczesnej technologii[59].

Jednocześnie literatura wskazuje na potencjalne stosowanie tzw. *zasady winy organizacyjnej*, której fundament stanowi standardowa zasada winy. Wskazany rodzaj obejmuje okoliczności, w których konkretne narzędzie/urządzenie funkcjonujące na podstawie sztucznej inteligencji nie podlegało adekwatnemu serwisowaniu i/lub nie aktualizowano w nim poprawek zgodnie z rekomendacjami twórcy[60].

Kodeks cywilny wyróżnia rodzaj odpowiedzialności opartej na zasadzie ryzyka, która zależna jest od zaistnienia zdarzenia, które zgodnie z ustawodawcą generuje tę odpowiedzialność. Zdarzenia te uwzględniają okoliczności o ponadstandardowym poziomie niebezpieczeństwa. Głównym założeniem tej odpowiedzialności jest jej niezależność od winy własnej osoby naruszającej, a jednocześnie brak przesłanek, że naprawienie szkody odbywałoby się na zasadach współżycia społecznego[61]. Zatem jeśli posługiwanie się konkretnym systemem wyposażonym w SI koreluje ze zwiększonym ryzykiem wystąpienia szkody, to odpowiedzialność spoczywająca na użytkowniku bądź twórcy systemu powinna być zaostrzona[62].

Ostatnią zasadą odpowiedzialności jest zasada opierająca się na słuszności. Zaistnieć może w przypadku, gdy norma prawna, która obliguje do naprawienia wyrządzonej szkody odnosi się do zasad współżycia społecznego[63].

Wskazuje się, że jednoznaczny wybór najbardziej adekwatnego rodzaju modelu odpowiedzialności jest zagadnieniem problematycznym. Zauważa się dychotomię, w której niektórzy autorzy optują za przyjmowaniem założeń odpowiedzialności opartej na zasadzie winy, podczas gdy inni postulują implementację innowacyjnych modeli łączących odpowiedzialność na zasadzie ryzyka z użytkowaniem nowoczesnych technologii[64].

---

[56] M. Serwach, *Wina jako zasada odpowiedzialności cywilnej oraz okoliczność zwalniająca z obowiązku naprawienia szkody*, „Wiadomości Ubezpieczeniowe" 1, 2009, s. 84.
[57] J. Gudowski, *Kodeks cywilny. Komentarz*, Warszawa 2013, s. 387; M. Sobas, *Cywilnoprawna odpowiedzialność za szkody wyrządzone przez zwierzęta na gruncie art. 431 Kodeksu Cywilnego*, „Roczniki Administracji i Prawa" 15(2), 2015, s. 151.
[58] Szerzej: R. Yampolskiy, *Unpredictability of AI: On the impossibility of accurately predicting all ac-tions of a smarter agent*, „Journal of Artificial Intelligence and Consciousness" 7(1), 2020, s. 109–118.
[59] P. Machnikowski, *Prawo deliktowe wobec nowych technologii*, Warszawa 2023, s. 177-183.
[60] P. Kaniewski, K. Kowacz, *Odpowiedzialność cywilna za szkody spowodowane funkcjonowaniem sztucznej inteligencji – uwagi de lege lata i de lege ferenda*, „Palestra" 11, 2024, s. 12.
[61] M. Moska, *Odpowiedzialność deliktowa za stwarzanie szczególnego niebezpieczeństwa dla otoczenia w XXI wieku a społeczne poczucie sprawiedliwości*, „Palestra" 1 2025, s. 57-83.
[62] P. Machnikowski, *op. cit*., s. 322-325.
[63] *Ibidem*, s. 220-226.
[64] J. Giezek, *op. cit*., s. 60-63.

ODPOWIEDŹ UNII EUROPEJSKIEJ NA ZAGROŻENIA ZWIĄZANE Z SI

Warta zauważenia jest również próba odpowiedzi ze strony Unii Europejskiej na pojawiające się niebezpieczeństwa wynikające z funkcjonowania oraz użytkowania technologii wykorzystujących SI. Istotnym dokumentem okazuje się Rozporządzenie Parlamentu Europejskiego i Rady (UE) nr 2024/1689 z dnia 13 czerwca 2024 r. w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji (akt w sprawie sztucznej inteligencji) - AI Act[65].

Ustawodawca, wprowadzając omawiane rozporządzenie miał na celu m.in. zapobieganie tego typu incydentom, które godziłyby w prawa człowieka. W AI Act wymienia się szereg zakazanych praktyk np. bezwzględny zakaz korzystania z systemów zdalnych w celu identyfikacji obywateli w sferze publicznej. Skoncentrowano się również na zagadnieniu niedoskonałości algorytmów SI[66]. W celu zminimalizowania szkodliwych skutków działania tychże algorytmów prawodawca wprowadził podział systemów SI według poziomu ryzyka[67]. Wymogiem odnoszącym się do funkcjonowania narzędzi sztucznej inteligencji jest zasada transparentności[68].

Ponadto Komisja Europejska zaproponowała projekt dyrektywy unijnej odnoszącej się do odpowiedzialności za SI – „AILD" (Dyrektywa Parlamentu Europejskiego i Rady w sprawie dostosowania przepisów dotyczących pozaumownej odpowiedzialności cywilnej do sztucznej inteligencji), której opracowanie skoncentrowano na wypełnieniu istniejącej luki w zakresie odpowiedzialności za szkody spowodowane działalnością systemów SI[69].

Wśród założeń AILD podkreślano konieczność implementacji zmian w zakresie istniejących uregulowań dotyczących czynów niedozwolonych popełnianych z udziałem sztucznej inteligencji. Zatem projekt ten skoncentrowany został w obszarze odpowiedzialności deliktowej. W kontekście działalności SI – na odpowiedzialności za wszelkie ryzyka i szkody wynikające z funkcjonowania sztucznej inteligencji. Dokument ten zakłada wiele zmian w dotychczasowych regulacjach. Głównym ułatwieniem jest możliwość żądania dostarczenia wszelkich informacji z zakresu działania systemu SI. Ponadto AILD wskazuje na zobowiązanie powoda do udowodnienia winy pozwanego oraz wykazania związku przyczynowo-skutkowego między systemem SI a wyrządzoną szkodą[70].

---

[65] 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), Dz.U.UE.L.2024.1689 z dnia 2024.07.12.

[66] K.Palacz, *Jak inteligentne jest prawo regulujące sztuczną inteligencję? Próba analizy AI Act na podstawie jej ugruntowania w przestrzeni komercyjnego użycia*, „Prawo Mediów Elektronicznych" 1, 2022, s. 23.

[67] P. Burczaniuk, „*Tworzenie prawa sztucznej inteligencji – wyzwania i perspektywy*, „Prawo i więź" 3(50), 2024, s. 287.

[68] D. Flisak, *Akt w sprawie sztucznej inteligencji*, sip.lex.pl/komentarze-i-publikacje/komentarze-praktyczne/akt-w-sprawie-sztucznej-inteligencji-470276293 (26.05.2026).

[69] M. Ziosi et al., *The EU AI Liability Directive (AILD): Bridging Information Gaps*, „European Journal of Law and Technology", 14(3), 2023, s. 2.

[70] *Ibidem*.

PODSUMOWANIE

Dokonana w toku niniejszej pracy wieloaspektowa i kompleksowa analiza – oparta na zaplanowanej triangulacji metodologicznej – umożliwiła dokonanie szczegółowej weryfikacji postawionej hipotezy badawczej. Zgromadzone dane dostarczyły przesłanek potwierdzających, iż algorytmy SI w swojej obecnej fazie rozwoju generują trudne do natychmiastowego rozpoznania ryzyko suicydogenne. Zagrożenie to manifestuje się w sposób szczególny w okolicznościach, w których jednostka wykazuje predyspozycje do dekompensacji emocjonalnej i/lub zmaga się z kryzysem psychicznym. Kwerenda dostępnych danych, obejmujący krytyczny wgląd w literaturę przedmiotu, najnowsze raporty, a także analizę studiów przypadków, prezentuje wysoce niepokojącą tendencję. Okazuje się, że interakcje między człowiekiem zmagającym się z trudnościami natury psychicznej a systemami opartymi o maszynowe uczenie wykazały wpływ o charakterze dyrektywnym na procesy decyzyjne użytkownika. Warunki zespołu presuicydalnego, cechujące się zawężeniem świadomości, w okolicznościach zjawisk tj. bańki informacyjne, sykofancja i złudne poczucie bliskości, mogą spotęgować, a także przyspieszyć transformacje z myśli rezygnacyjnych i samobójczych do konkretnych aktów autodestrukcji. Dodatkowo przeprowadzone badania dowiodły braki w zakresie implementacji wystarczających procedur bezpieczeństwa w najpopularniejszych dostępnych powszechnie modelach SI. Ocena uwarunkowań polskiego Kodeksu karnego oraz Kodeksu cywilnego, w zestawieniu z najnowszymi regulacjami Unii Europejskiej ujawnia luki ustawodawcze. Brak jednoznacznych regulacji w kontekście odpowiedzialności za wyrządzone szkody uniemożliwia skuteczną ochronę bezpieczeństwa grup społecznych.

BIBLIOGRAFIA

Adamkiewicz Marek. 2015. „Suicydologiczny wizerunek niebezpieczeństwa". Studia Bezpieczeństwa Narodowego 5(7): 225–252. https://doi.org/10.37055/sbn/135293.

Bathaee Yavar. 2018. "The artificial intelligence black box and teh failure of intent and causation". Springer 31(2): 890–938.

Bhuiyan Johana. 2025. ChatGPT encouraged Adam Raine's suicidal thoughts. His family's lawyer says OpenAI knew it was broken. In www.theguardian.com/us-news/2025/aug/29/chatgpt-suicide-openai-sam-altman-adam-raine.

Blandyna Lewkowicz Monika. 2024. 14-latek zakochał się w chatbocie i popełnił samobójstwo". In cyberdefence24.pl/social-media/14-latek-zakochal-sie-w-chatbocie-i-popelnil-samobojstwo-jest-pozew.

Boppiniti Sai T. 2022. „Ethical implications of srtifical intelligence: a review od early research and perspectives", International engineering Journal for Research & Development 7(1): 1- 8.

Burczaniuk Piotr. 2024. „Tworzenie prawa sztucznej inteligencji – wyzwania i perspektywy", Prawo i więź 3 (50): 283-300. https://doi.org/10.36128/PRIW.VI51.707.

Denecke Kerstin, Lopez-Campos Guillermo, Gabarron Elia . 2025. „Hidden in plain sight: the harmful side of AI–based mental health interventions In Elisavet Andrikopoulou, Parisis

Gallos, Theodoros N. Arvanitis, Rosalynn Austin, Arriel Benis (eds), Intelligent health systems – from technology to data and knowledge, 248-252. IOS Press.

Dobson James E. 2023. "On reading and interpreting black box deep neural networks". Springer 5 (2023): 431–449. https://link.springer.com/article/10.1007/s42803-023-00075-w.

Flisak Damian. 2023. Akt w sprawie sztucznej inteligencji. In sip.lex.pl/komentarze-i-publikacje/komentarze-praktyczne/akt-w-sprawie-sztucznej-inteligencji-470276293.

Fontenelle Paula. 2025. ChatGPT Made Him Delusional. A story about AI, loneliness, shame, and being human. In www.psychologytoday.com/us/blog/understanding-suicide/202511/chatgpt-made-him-delusional.

Formella Zbigniew. 2004. „Samobójstwo. Refleksja psychopedagogiczna", Seminare 20: 369–385. https://doi.org/10.21852/sem.2004.27.

Gazeta Prawna, Chatbot namówił mężczyznę do popełnienia samobójstwa. In www.gazetaprawna.pl/wiadomosci/swiat/artykuly/8690927,belgia-media-chatbot-namowil-mezczyzne-do-popelnienia-samobojstwa.html.

Giezek Jacek. 2024. „Sztuczna inteligencja a fundamenty odpowiedzialności karnej państwa demokratycznego". Studia nad Autorytaryzmem i Totalitaryzmem 46(4): 55-72. https://doi.org/10.19195/2300-7249.46.4.4.

Grudecki Michał. 2024. „Prawnokarna ocena prób samobójczych". Prawo i więź 3(50): 329-359. https://doi.org/10.36128/PRIW.VI50.786.

Gudowski Jacek. 2013. Kodeks cywilny. Komentarz. Warszawa: LexisNexis Polska.

Head Keith Rober. 2026. „Digital companions, real casualties: A commentary on rising AI-related mental health crises". Current Research in Psychiatry 6(1): 1-5. https://doi.org/10.46439/Psychiatry.6.043

Hill Kashmir. 2025. Lawsuits blame ChatGPT for suicides and harmful delusions. The New York Times. In www.nytimes.com/2025/11/06/technology/chatgpt-lawsuit-suicides-delusions.html.

Hołyst Brunon.2002. *Suicydologia*. Warszawa: LexisNexis.

Hołyst Brunon2024. *Suicydologia*. Warszawa: Wolters Kluwer

Idzik Jakub, Klepka Rafał. 2021. „Bańka informacyjna i zjawiskowa komora echa". W Olga-Wasiuta, Sergiusz Wasiuta (red.), Encyklopedia bezpieczeństwa, 201–209. Wydawnictwo Libron.

Jankowska Marlena. 2015. „Podmiotowość prawna sztucznej inteligencji?". W A. Bielska-Brodziak (red.), O czym mówią prawnicy, mówiąc o podmiotowości, 171-196. Wydawnictwo Uniwersytetu Śląskiego.

Jarosz Maria. 2018. „Samobójstwa". Civitas. Studia z filozofii polityki 22: 215–226. https://doi.org/10.35757/CIV.2018.22.08.

Kaniewski Piotr, Kowacz Krzysztof. 2024. „Odpowiedzialność cywilna za szkody spowodowane funkcjonowaniem sztucznej inteligencji – uwagi de lege lata i de lege ferenda". Palestra 11: 6-33. https://doi.org/10.54383/0031-0344.2024.11.2

Karamali Kamil. 2025. 'A world-saving mission': Ontario man alleges ChatGPT drove him to psychosis, CTV News. In www.ctvnews.ca/canada/article/ontario-man-alleges-chatgpt-caused-delusions-sues-parent-company-openai/.

Kielan Aleksandra, Bąbik Katarzyna, Cieślak Ilona, Dobaczewska Paula. 2017. „Religia katolicka z zachowania suicydalne w Polsce". Medycyna Ogólna i Nauki o Zdrowiu 23(2): 158–164. https://doi.org/10.26444/monz/74271.

Kooli Chokri. 2025. „Generative Artificial intelligence Addiction Syndrome: A New Behavioral Disorder?". Asian Journal of Psychiatry 107(1). https://doi.org/10.1016/j.ajp.2025.104476.

Krasuski Andrzej. 2021. Status prawny sztucznego agenta. Podstawy prawne zastosowania sztucznej inteligencji. Warszawa: C. H. Beck.

Kubera Paulina. 2026. „Odpowiedzialność za szkody wynikające z działania systemów sztucznej inteligencji". Ruch prawniczy, ekonomiczny i socjologiczny 1: 89-91. https://doi.org/10.14746/rpeis.2026.88.1.05

Machnikowski Piotr. 2023. Prawo deliktowe wobec nowych technologii. Warszawa: Wolters Kluwer.

Małecki Mikołaj. 2021. „Współudział w samobójstwie (kwestia dobra pranego)". Acta Iuris Stetinensis 3(35): 53–63. https://doi.org/10.18276/ais.2021.35-04 .

Marshall Brandeis. 2023. „No legal personhood for AI". Patterns 4(11). https://doi.org/10.1016/j.patter.2023.100861.

Maslej Nestor et al., Artificial intelligence index report 2025, Stanford Institute for Human-Centered Artificial Intelligence. In arxiv.org/pdf/2504.07139.

Mladenovic Miljana, Ošmjanski Vera, Stanković Stasa V. 2021. „Cyber-aggression, Cyberbullying, and Cyber-grooming: A Survey and Research Challenges", ACM Computing Surveys 54(1): 1-11. https://doi.org/10.1145/3424246.

Moska Monika. 2025. „Odpowiedzialność deliktowa za stwarzanie szczególnego niebezpieczeństwa dla otoczenia w XXI wieku a społeczne poczucie sprawiedliwości". Palestra 1: 57-83. https://doi.org/10.54383/0031-0344.2025.01.6

Oleksiewicz Izabela. 2022. "Artificial intelligence versus human – a threat or a necessity of evolution". European Studies Quarterly 3: 55–69. https://doi.org/10.31338/1641-2478pe.3.22.4.

Ortutay Barbara and The Associated Press. 2025. He was 17 and asked ChatGPT for help. It allegedly told him how to die instead. In fortune.com/2025/11/07/chatgpt-open-sam-altman-suicide-lawsuits-17-year-old/.

Palacz Kamil. 2022. „Jak inteligentne jest prawo regulujące sztuczną inteligencję? Próba analizy AI Act na podstawie jej ugruntowania w przestrzeni komercyjnego użycia". Prawo Mediów Elektronicznych: 22-25.

Poenaru Liviu. 2025. AI psychosis and the "AI neurosis": Chatbot-induced delusions and neurotic decompensation. E.U.LABORATORY. In https://www.eulaboratory.com/ai-psychosis-and-the-ai-neurosis.

Prochowicz Katarzyna. 2009. „Obłęd we dwoje. Objawy i psychospołeczne uwarunkowania indukowanych zaburzeń urojeniowych". Psychiatria Polska 43(1): 19–30.

Rejmaniak Rafał. 2021.„Autonomiczność systemów sztucznej inteligencji jako wyzwanie dla prawa karnego". Roczniki Nauk Prawnych 31(3): 95–113. https://doi.org/10.18290/rnp21313.6

Rezner Beata, Witkowska Małgorzata. 2024. „Wyzwania i niebezpieczeństwa wynikające z wykorzystywania sztucznej inteligencji w procesie podejmowania decyzji", Zeszyty Naukowe. Polski Uniwersytet na obczyźnie w Londynie 12: 169-180.

Sapota Tomasz. 2009. „Rzymska idea samobójstwa". Symbolae philologorum posnaniensium Graecae et latinae 19: 281-287.

Seed Ahmed, Javaid Tali. 2025. Sycophancy in AI: Challenges in Large Language Models and Argumentation Graphs. In www.researchgate.net/publication/389939533_Sycophancy_in_AI_Challenges_in_Large_Language_Models_and_Argumentation_Graphs.

Serwach Małgorzata. 2009. „Wina jako zasada odpowiedzialności cywilnej oraz okoliczność zwalniająca z obowiązku naprawienia szkody". Wiadomości Ubezpieczeniowe 1: 84-96. https://doi.org/10.13140/RG.2.2.14094.06723.

Sobas Magdalena. 2015. „Cywilnoprawna odpowiedzialność za szkody wyrządzone przez zwierzęta na gruncie art. 431 Kodeksu Cywilnego". Roczniki Administracji i Prawa 15(2): 149-167. https://doi.org/10.13140/RG.2.2.14094.06723.

Stokel-Walker Chris. 2025. „AI driven psychosis and suicide are on the rise, but what happens if we turn the chatbots off?". The BJM 391. https://doi.org/10.1136/bmj.r2239.

Świerczyński Marek. 2020. „Sztuczna inteligencja w prawie prywatnym międzynarodowym – wstępne rozważania". Problemy Prawa Prywatnego Międzynarodowego 25: 27–42. https://doi.org/10.13140/RG.2.2.14094.06723.

Tekliński Jarosław. 2018. „Czy człowiek ma prawo do samobójstwa?". Humanistica 21(2): 97–128.

Tomasz Zalewski. 2020. Prawo sztucznej inteligencji. Warszawa: C.H. Beck.

Udalova Larisa V. 2024. „Suicidal Behavior as an Indicator of Social Pathology". Contemporary Philosophical Research. https://doi.org/10.18384/2949-5148-2023-4-113-119.

Vink Ton 2024. „David Hume on Suicide and the Value of Human Life: A European Legacy. The European Legacy 29 (7–8): 748–766. https://doi.org/10.1080/10848770.2024.2385744.

Wikipedia. Raine v. OpenAI. In en.wikipedia.org/wiki/Raine_v._OpenAI.

Wilczyńska Olga. 2008. „Metafizyka śmierci w prozie Olgi Tokarczuk". Acta Universitatis Lodziensis. Folia Litteraria Polonica 11: 183–186. https://doi.org/10.18778/1505-9057.11.13.

World Health Organization. 2025. WHO global health estimates: Suicide data. In www.who.int/teams/mental-health-and-substance-use/data-research/suicide-data.

Yampolskiy Roman.2020. „Unpredictability of AI: On the impossibility of accurately predicting all ac-tions of a smarter agent". Journal of Artificial Intelligence and Consciousness 7(1): 109–118. https://doi.org/10.1142/S2705078520500034

Yeung Joshus A., Dalmasso Jacopo, Foshini Luca, Dobson Richard JB., Kraljevic Zeljko. 2025. The Psychogenic Machine: Simulating AI Psychosis, Delusion Reinforcement and Harm Enablement in Large Language Models. University College London. arxiv.org/abs/2509.10970.

Yousaf Nadine. 2025. BBC. Parents of teenager who took his own life sue OpenAI. *BBC. In* www.bbc.com/news/articles/cgerwp7rdlvo.

Zero Daniel. 2021. „Cybersuicydologia – nowe technologie a samobójstwo". Kortowski Przegląd Prawniczy 1: 35–53. https://doi.org/10.31648/kpp.6688.

Ziosi Marta, Mökander Jakob, Novelli Claudio, Casolari Federico, Taddeo Mariarosaria, Floridi Luciano. 2023. „The EU AI Liability Directive (AILD): Bridging Information Gaps". European Journal of Law and Technology 14(3): 1-10. https://dx.doi.org/10.2139/ssrn.4470725.