

RAFAŁ REJMANIAK
Uniwersytet w Białymstoku
r.rejmaniak@uwb.edu.pl
ORCID: <https://orcid.org/0000-0003-1908-5844>

AUTONOMICZNOŚĆ SYSTEMÓW SZTUCZNEJ INTELIGENCJI JAKO WYZWANIE DLA PRAWA KARNEGO

AUTONOMY OF AI SYSTEMS AS A CHALLENGE FOR CRIMINAL LAW

Abstract: Challenges of a previously unknown nature. One of the key problems in criminal law related to the dissemination of artificial intelligence systems is determining the entity responsible for possible violations of legally protected goods caused by such systems. The aim of the study is an attempt to answer the question: how the increase in the level of autonomy of artificial intelligence systems affects the possibility of assigning criminal liability under Polish law for the effects caused by the actions of such systems. The article will analyze the challenges posed by AI systems with varying degrees of autonomy and will examine the possibilities of using the guarantor's concept as the entity responsible for harm caused by AI system.

Keywords: artificial intelligence; guarantor; autonomy; AI system

WPROWADZENIE

Ciągły postęp w dziedzinie Sztucznej Inteligencji (SI) generuje coraz trudniejsze do rozwiązania problemy dotyczące określenia ram prawnych i moralnych jej tworzenia oraz wykorzystywania. Systemy SI charakteryzują się bowiem postępującą autonomicznością – ich funkcjonowanie jest coraz mniej zależne od kontrolującego je człowieka. Sytuacja taka stawia nowe wyzwania zarówno o charakterze moralnym, jak i prawnym. Wydaje się, że obecnie jednym z węzłowych problemów w tym zakresie jest określenie zasad odpowiedzialności za potencjalne szkody, które mogą zostać wyrządzone w dobrach prawnych w wyniku działania systemów SI. W niniejszym artykule podjęto próbę zarysowania tych problemów, ograniczając się jedy-

nie do kwestii związanych z przypisaniem odpowiedzialności karnej. Pierwsza część opracowania poświęcona zostanie omówieniu pojęcia sztucznej inteligencji, następnie zaś analizie poddane zostaną wyzwania niesione dla prawa karnego, które kształtują się odmiennie w zależności od stopnia autonomiczności systemu SI.

1. POJĘCIE SZTUCZNEJ INTELIGENCJI

Wprowadzenie pojęcia „sztuczna inteligencja” do naukowej dyskusji przypisuje się Johnowi McCarthy’emu¹. W pierwotnym rozumieniu pojęcie to miało oznaczać „[proces, który] sprawia, że maszyna zachowuje się w sposób, który nazwalibyśmy inteligentnym, gdyby w ten sposób zachowywał się człowiek”². Jeżeli więc obiekt taki, z perspektywy zewnętrznego obserwatora, wystarczająco dobrze imitowałby zachowania człowieka, zyskałby miano sztucznej inteligencji³.

Postęp technologiczny, który dokonał się od lat 50. XX wieku, skutkował jednak zmianą rozumienia pojęcia sztucznej inteligencji. Ambitne założenia o skonstruowaniu sztucznego umysłu – analogicznego do umysłu człowieka – okazały się trudniejsze do zrealizowania niż pierwotnie zakładano. Choć stworzenie świadomej sztucznej inteligencji (nazywanej wcześniej „silną SI”, a obecnie niekiedy „ogólną SI”)⁴ stanowi „święty Graal” badaczy, to nie wiadomo, kiedy i czy w ogóle będzie to możliwe⁵. Współczesne systemy SI zdolne są jedynie do realizowania konkretnych wyznaczonych im celów, zaliczają się więc do kategorii „wąskiej SI”.

¹ M. TEGMARK, *Życie 3.0. Człowiek w erze sztucznej inteligencji*, Warszawa: Prószyński i S-ka 2019, s. 60.

² J. MCCARTHY, M.L. MINSKY, N. ROCHESTER, C.E. SHANNON, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, 1955, <http://www-formal.stanford.edu/jmc/history/dartmouth.pdf> [dostęp: 6.02.2021].

³ Zob. A.M. TURING, *Computing Machinery and Intelligence*, „Mind” 59 (1950), nr 236, s. 433-436.

⁴ M.A. BODEN, *Sztuczna inteligencja. Jej natura i przyszłość*, Łódź: Wydawnictwo Uniwersytetu Łódzkiego 2020, s. 33-34.

⁵ Potencjalny moment uzyskania przez system SI świadomości nazywany jest „osobliwością technologiczną” – zob. V. VINGE, *The Coming Technological Singularity: How to Survive in the Post-Human Era*, [w:] *Vision-21: Interdisciplinary Science and Engineering in the Era of Cyberspace*, red. G. Landis, S. Bailey, Ohio: NASA Conference Publication 10129, NASA Lewis Research Center 1993, s. 12.

Z jednej strony wykorzystywanie systemów SI w wielu sferach rzeczywistości społecznej, np. w służbie zdrowia, w przemyśle, w usługach, w sektorze publicznym czy w energetyce, usprawnia pracę i przynosi wielkie korzyści. Z drugiej jednak, dostrzegalne są też zagrożenia niesione przez upowszechnianie się SI⁶. Problematyka ta wzbudza coraz większe zainteresowanie wielu państw, jak i żywo dyskutowana jest na poziomie międzynarodowym.

Mimo istnienia wielu definicji SI, wydaje się zasadne odniesienie się do tej zaproponowanej przez Grupę Ekspertów Wysokiego Szczebla ds. Sztucznej Inteligencji (powołaną przez Komisję Europejską w 2018 r.), bowiem to Unia Europejska jest podmiotem, który obecnie wyznacza standardy regionalne w dziedzinie SI. W myśl tej definicji systemy SI to:

oprogramowania komputerowe (i ewentualnie również sprzęt komputerowy) stworzone przez człowieka, które, biorąc pod uwagę założony cel, działają w wymiarze fizycznym lub cyfrowym poprzez postrzeganie ich otoczenia dzięki gromadzeniu danych, interpretacji zebranych ustrukturyzowanych lub nieustrukturyzowanych danych, rozumowaniu na podstawie wiedzy lub przetwarzaniu informacji pochodzących z tych danych oraz podejmowaniu decyzji w sprawie najlepszych działań, które należy podjąć w celu osiągnięcia określonego celu. Systemy SI mogą wykorzystywać symboliczne reguły albo uczyć się modelu numerycznego, a także dostosowywać swoje zachowanie, analizując wpływ ich poprzednich działań na otoczenie⁷.

Coraz częściej jednak formułuje się definicje eksponujące interakcję takich systemów ze środowiskiem, w którym się znajdują. We Wniosku Komisji Europejskiej odnoszących się do propozycji Rozporządzenia Parlamentu Europejskiego i Rady ustanawiającego zharmonizowane przepisy dotyczące sztucznej inteligencji (Akt w sprawie sztucznej inteligencji)⁸, system sztucznej inteligencji definiuje się jako „oprogramowanie opracowane przy użyciu co najmniej jednej spośród technik i podejść wymienionych w załączniku I, które może – dla danego zestawu celów określonych przez człowieka – ge-

⁶ W. FILIPKOWSKI, *Prawo karne wobec sztucznej inteligencji*, [w:] *Prawo sztucznej inteligencji*, red. L. Lai, M. Świerczyński, Warszawa: Wydawnictwo C.H. Beck 2020, s. 124-125.

⁷ *Definicja SI: Główne funkcje i dyscypliny*, Grupa ekspertów wysokiego szczebla ds. sztucznej inteligencji, Bruksela: Komisja Europejska 2019, s. 6, https://www.europarl.europa.eu/meet/docs/2014_2019/plmrep/COMMITTEES/JURI/DV/2019/11-06/Ethics-guidelines-AI_PL.pdf [dostęp: 6.02.2021].

⁸ *Wniosek – Rozporządzenie Parlamentu Europejskiego i Rady ustanawiające zharmonizowane przepisy dotyczące sztucznej inteligencji (Akt w sprawie sztucznej inteligencji) i zmieniające niektóre inne akty ustawodawcze Unii*, {SEC(2021) 167 final} – {SWD(2021) 84 final} – {SWD(2021) 85 final} COM(2021) 206 final, 2021/0106 (COD), Bruksela: Komisja Europejska 2021, s. 47, https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.00012.02/DOC_1&format=PDF [dostęp: 13.07.2021].

nerować wyniki, takie jak treści, przewidywania, zalecenia lub decyzje wpływające na środowiska, z którymi wchodzi w interakcję”. W załączniku do wniosku wskazano następujące techniki i podejścia:

a) mechanizmy uczenia maszynowego, w tym uczenie nadzorowane, uczenie się maszyn bez nadzoru i uczenie przez wzmacnianie, z wykorzystaniem szerokiej gamy metod, w tym uczenia głębokiego;

b) metody oparte na logice i wiedzy, w tym reprezentacja wiedzy, indukcyjne programowanie (logiczne), bazy wiedzy, silniki inferencyjne i dedukcyjne, rozumowanie (symboliczne) i systemy ekspertowe;

c) podejścia statystyczne, estymacja bayesowska, metody wyszukiwania i optymalizacji.

Autonomiczność systemów SI należy rozumieć jako zdolność do samodzielnego działania bez ingerencji człowieka. Jest to cecha stopniowalna, wyznaczana na osi pomiędzy całkowitym jej brakiem (system jest narzędziem w ręku człowieka i nie podejmuje żadnych decyzji) a pełną autonomicznością (system działa bez jakiegokolwiek ingerencji człowieka). W literaturze podejmuje się próby konstruowania stosownych skal, pozwalających na określenie stopnia autonomiczności systemu i zakresu ingerencji człowieka. Poniżej zaprezentowano zestawienie trzech popularnych skal: Skalę T. Sheridana określającą poziomy interakcji z systemem automatycznym, skalę SAE J3016 określającą poziomy autonomiczności pojazdów i skalę opartą o pętlę decyzyjną stosowaną do analizy autonomiczności systemów o charakterze militarnym.

Tabela 1. Skale autonomiczności

Skala T. Sheridana	SAE J3016 Autonomy Levels	Pętla decyzyjna
1. Człowiek podejmuje wszystkie decyzje i działania, system nie oferuje żadnego wsparcia.	1. Wszystkie czynności związane z prowadzeniem wykonywane są przez człowieka.	1. <i>Man-in-the-loop</i> – system nie ma żadnej swobody działania. Człowiek steruje systemem i podejmuje decyzje lub je zatwierdza przed wykonaniem.
2. System proponuje zespół alternatywnych działań do wyboru przez człowieka, albo...	2. Wszystkie czynności związane z prowadzeniem wykonywane są przez człowieka z asystą systemu, który wykorzystuje informacje z otoczenia, lecz to człowiek odpowiedzialny jest za kontrolę podczas zmieniających się warunków.	
3. System zawęża propozycje do jedynie kilku działań, albo...		
4. System proponuje jedno rozwiązanie.		

	3. Wszystkie czynności związane z prowadzeniem wykonywane są przez człowieka z asystą jednego lub kilku systemów, który wykorzystuje informacje z otoczenia, lecz to człowiek odpowiedzialny jest za kontrolę podczas zmieniających się warunków.	
5. System wykonuje proponowane działanie po akceptacji człowieka.	4. Wszystkie czynności wykonywane są przez system podczas całego procesu prowadzenia pojazdu. Kierowca jest jednak w ciągłej gotowości, aby przejąć sterowanie nad pojazdem.	2. <i>Man-on-the-loop</i> – system podejmuje i wykonuje decyzje bez aktywnej roli człowieka. Człowiek nie bierze udziału w procesie decyzyjnym, ale sprawuje nadzór nad systemem i może przerwać jego działanie (<i>veto</i>), gdy uzna to za konieczne.
6. System samodzielnie decyduje o wykonaniu działania, człowiek może przerwać działanie w ograniczonym czasie przed podjęciem działania.		
7. System samodzielnie wykonuje działanie, a następnie informuje o tym człowieka.	5. Wszystkie czynności wykonywane są przez system podczas całego procesu prowadzenia pojazdu, nawet jeśli człowiek nieodpowiednio reaguje na prośby o interwencję.	3. <i>Man-out of-the-loop</i> – system działa w sposób niezależny od człowieka.
8. System samodzielnie wykonuje działanie, a następnie informuje o tym człowieka, jeżeli człowiek zażąda informacji.		
9. System samodzielnie wykonuje działanie, a następnie informuje o tym człowieka, jeżeli uzna to za konieczne.		
10. System działa w pełni autonomicznie, ignorując człowieka.	6. Wszystkie czynności wykonywane są przez system podczas całego procesu prowadzenia pojazdu. System przejmuje kontrolę w każdej sytuacji, na każdej drodze, w każdych warunkach.	

Źródło: Opracowanie własne⁹

⁹ Na podstawie R. PARASURAMAN, T.B. SHERIDAN, CH.D. WICKENS, *A Model for Types and Levels of Human Interaction with Automation*, „IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans” 30 (2000), nr 3, s. 287; A. LELE, *Debating Lethal Auto-*

Systemy SI w swej istocie są zbiorami programów (oprogramowaniem) lub całymi złożonymi systemami działania. Rozwiązania, na których oparte jest funkcjonowanie takich systemów, od „zwykłych” algorytmów, rozumianych jako realizacja sekwencji z góry określonych kroków, odróżnia zdolność do „uczenia się”. Uczenie takich systemów, nazywane uczeniem maszynowym, rozumiane jest jako wykorzystywanie zewnętrznych danych empirycznych w celu tworzenia i aktualizacji podstaw dla udoskonalonego działania na podobnych danych w przyszłości oraz wyrażania tych podstaw w zrozumiałej i symbolicznej postaci¹⁰. Nie jest celem niniejszego opracowania prezentowanie technik uczenia maszynowego¹¹, jednak z perspektywy dalszych rozważań dotyczących kwestii przypisania odpowiedzialności za skutki powodowane przez systemy SI jako systemy samouczące, niezbędne jest poczynienie pewnej uwagi.

Jedną ze współcześnie najbardziej popularnych technik jest uczenie głębokie oparte o wiele warstw sztucznych sieci neuronowych. Sztuczna sieć neuronowa jest uproszczonym matematycznym modelem struktury mózgu¹². Sztuczne neurony połączone w taką sieć otrzymują sygnały wejściowe, a każdy taki sygnał mnożony jest przez odpowiednią wartość liczbową (wagę), może ona być pobudzająca albo opóźniająca. Jeżeli zostanie przekroczony próg aktywacji, neuron wysyła sygnał, który stanowi sygnał wejściowy dla neuronów w kolejnej warstwie¹³. W tym przypadku uczenie polega na ustalaniu odpowiednich wag dla różnych sygnałów wejściowych. Istotnym problemem związanym z uczeniem głębokim jest jednak fakt, że pomiędzy warstwą wejściową i wyjściową istnieją jeszcze warstwy ukryte, a same sieci nie posiadają możliwości wyjaśniania procesu podejmowania decyzji¹⁴. O ile jeszcze możliwe jest ustalenie, jakie wagi system przypisał poszczególnym

nomous Weapons Systems, „Journal of Defense Studies” 13 (2019), nr 1, s. 55-56; T. NEUMANN, *Perspektywy wykorzystania pojazdów autonomicznych w transporcie drogowym w Polsce*, „Autobusy” 2018, nr 12, s. 788.

¹⁰ D. MICHIE, *Methodologies from Machine Learning in Data Analysis and Software*, „The Computer Journal” 34 (1991), nr 6, s. 562.

¹¹ Szczegółowe informacje można znaleźć np. w: K. RÓŻANOWSKI, *Sztuczna inteligencja: rozwój, szanse i zagrożenia*, „Zeszyty Naukowe Warszawskiej Wyższej Szkoły Informatyki” 2007, nr 2; M. FLASIŃSKI, *Wstęp do sztucznej inteligencji*, Warszawa: Wydawnictwo Naukowe PWN 2020; L. RUTKOWSKI, *Metody i techniki sztucznej inteligencji*, Warszawa: Wydawnictwo Naukowe PWN 2012; M.A. BODEN, *Sztuczna inteligencja*, s. 59-60.

¹² M. FLASIŃSKI, *Wstęp do sztucznej inteligencji*, s. 161.

¹³ A. KASPERSKA, *Problemy zastosowania sztucznych sieci neuronalnych w praktyce prawniczej*, „Przegląd Prawa Publicznego” 2017, nr 11, s. 25.

¹⁴ Tamże, s. 27.

sygnałom wejściowym i ponowne przeszkolenie w przypadku niesatysfakcjonującego wyniku¹⁵, to nie jest już możliwe ustalenie, dlaczego system określił wagi właśnie w taki sposób. W rzeczywistości może się okazać, że wytrenowany system SI podejmować będzie oczekiwane decyzje w sposób poprawny, ale opierając się na zupełnie innych kryteriach, niż zakładali jego twórcy¹⁶, bądź też zachowa się w sposób całkowicie inny niż oczekiwany¹⁷. W przypadku niektórych systemów SI wprowadzenie ich do użytku następuje po ukończeniu treningu, inne natomiast uczą się przez cały czas użytkowania (np. Translator Google¹⁸).

2. SYSTEM SI JAKO NARZĘDZIE W RĘKACH CZŁOWIEKA

Przechodząc do rozważań dotyczących możliwości przypisania odpowiedzialności karnej za szkody wyrządzone w dobrach prawnie chronionych działaniem systemów SI, wyodrębnić należy trzy rodzaje sytuacji, w zależności od stopnia autonomiczności takiego systemu. Do analizy wykorzystana zostanie koncepcja pętli decyzyjnej, która, choć jest modelem uproszczonym, pozwoli na zarysowanie węzłowych problemów.

Przed rozpoczęciem analizy wskazanych powyżej typów sytuacji, należy zasygnalizować istotną kwestię ogólniejszej natury. Ze względu na to, że współcześnie istnieją techniczne rozwiązania pozwalające na monitorowanie zachowań człowieka w cyberprzestrzeni, istotnym zagadnieniem staje się możliwość zniekształcania lub zakłócania poprawnego działania systemu SI przez podmioty niebędące do tego uprawnionymi. W przypadku działania osoby trzeciej, która nie jest ani twórcą, ani uprawnionym użytkownikiem, autonomiczność systemu SI może być pewną iluzją. System bowiem zostaje zmanipulowany, aby działać w inny sposób, niż został do tego zaprojektowany (np. poprzez celowe zanieczyszczanie danych, w przypadku systemów uczących się w trakcie ich użytkowania). W takich przypadkach wydaje się zasadne traktowanie systemu SI jako narzędzia w rękach sprawcy, niezależnie

¹⁵ M. FLASIŃSKI, *Wstęp do sztucznej inteligencji*, s. 163.

¹⁶ Zob. M.T. RIBEIRO, S. SINGH, C. GUESTRIN, „Why Should I Trust You?” *Explaining the Predictions of Any Classifier*, San Francisco: The 22nd ACM SIGKDD International Conference 2016, s. 8-10, <https://www.kdd.org/kdd2016/papers/files/rfp0573-ribeiroA.pdf> [dostęp: 6.02.2021].

¹⁷ Zob. G. NEFF, P. NAGY, *Talking to Bots: Symbiotic Agency and the Case of Tay*, „International Journal of Communication” 2016, nr 10, s. 4920-4922.

¹⁸ G. MASSEY, M. EHRENSBERGER-DOW, *Machine learning: Implications for translator education*, „Lebende Sprachen” 62 (2017), nr 2, s. 301.

od tego, jaki poziom autonomiczności system w założeniu miał prezentować. Istotną rolę może jednak odegrać przyczynienie się podmiotu uprawnionego. Systemy SI coraz częściej funkcjonujące w ścisłym związku z innymi obiektami albo ludźmi (np. w ramach Internetu Rzeczy – IoT, czy Internetu Przemysłowego – IioT) są widoczne w środowisku, w którym działają. Co za tym idzie, są dostrzegalne również dla cyberprzestępców. Stąd też obowiązkiem twórcy staje się nie tylko zagwarantowanie prawidłowego funkcjonowania systemu „od wewnątrz”, ale również zapewnienie jego odporności na ataki „z zewnątrz”. Użytkownik natomiast zobligowany jest do dochowania ustalonych przez projektanta reguł wykorzystania danego systemu. Zignorowanie tego obowiązku może umożliwić lub ułatwić ingerencję w system SI ze strony nieuprawnionej osoby trzeciej.

Przechodząc do szczegółowych rozważań, w pierwszej kolejności omówiony zostanie typ sytuacji, które charakteryzują się brakiem swobody działania systemu SI i całkowitym uzależnieniem od jego operatora albo użytkownika. System taki może podejmować decyzje na podstawie analizy danych wejściowych proponując jedno bądź wiele działań, ostatecznie jednak to człowiek jest instancją decydującą o wyborze i realizacji któregoś z nich. Przykładem tego typu współpracy człowieka z systemem SI może być obszar diagnostyki w opiece zdrowotnej¹⁹. System SI może gromadzić i przetwarzać dane na przykład w postaci publikacji medycznych albo historii chorób wielu pacjentów, aby zaproponować diagnozę. Ocena, czy diagnoza jest trafna i decyzja, czy podjąć określone leczenie, spoczywać będzie natomiast na człowieku. W takim przypadku wydaje się nie budzić kontrowersji twierdzenie, iż to człowiek, który decyduje o wdrożeniu leczenia, ponosić będzie odpowiedzialność za ewentualny błąd w sztuce, skutkujący naruszeniem takich dóbr pacjenta, jak życie czy zdrowie²⁰. Niemniej jednak w praktyce pojawić się mogą problemy polegające na pokusie traktowania takiego systemu jako nieomylnego, swoistego „moralnego buforu”²¹ rzekomo chroniącego przed odpowiedzialnością użytkownika, niezdolnego przecież do przetwarzania tak wielkich zbiorów danych.

Aby ograniczyć zagrożenie pokładania zbyt dużego zaufania do decyzji systemu SI przez jego operatorów i użytkowników, wypracowuje się standardy

¹⁹ T. DAVENPORT, R. KALAKOTA, *The potential for artificial intelligence in healthcare*, „Future Healthcare Journal” 6 (2019), nr 2, s. 95-96.

²⁰ Podobnie na gruncie prawa cywilnego: L. BOSEK, *Perspektywy rozwoju odpowiedzialności cywilnej za inteligentne roboty*, „Forum Prawnicze” 52 (2019), nr 2, s. 13.

²¹ M.L. CUMMINGS, *Automation and Accountability in Decision Support System Interface Design*, „The Journal of Technology Studies” 32 (2006), nr 1, s. 26.

budowy i użytkowania takich systemów. Przykładowo, Biała Księga w sprawie sztucznej inteligencji wydana przez Komisję Europejską w dniu 19 lutego 2020 r.²² formułuje wymaganie, aby systemy sztucznej inteligencji były solidne i dokładne lub przynajmniej aby prawidłowo odzwierciedlały poziom dokładności na wszystkich etapach ich cyklu życia. Pojawia się ono również w innych dokumentach kształtujących standardy budowania systemów SI zarówno na gruncie poszczególnych państw, jak i na poziomie międzynarodowym²³.

3. SYSTEM SI NADZOROWANY PRZEZ CZŁOWIEKA

Jako znacznie bardziej problematyczny z perspektywy przypisania odpowiedzialności karnej jawi się kolejny typ sytuacji, w których system SI działa samodzielnie, bez aktywnego udziału człowieka. Człowiek natomiast przyjmuje rolę nadzorującego, władnego do przerwania działania systemu, na przykład poprzez jego wyłączenie, wstrzymanie wykonania decyzji czy przejęcie sterowania. O ile jeszcze we wcześniejszym przypadku możliwe jest dostrzeżenie aktywnej roli człowieka i powiązania jego odpowiedzialności karnej właśnie z tą rolą, o tyle w omawianym typie do szkody, wynikającej z działania systemu SI, może dojść bez udziału człowieka. Przykładem takiej sytuacji może być korzystanie z pojazdu autonomicznego, kiedy osoba znajdująca się w pojeździe zobowiązana jest do obserwowania trasy i w sytuacji krytycznej przejęcia sterowania.

Problem z ustaleniem podmiotu odpowiedzialnego za skutek spowodowany przez taki system jest wielopłaszczyznowy. Sam proces budowania i wdrażania systemu SI może angażować wiele podmiotów, takich jak producent, programiści opracowujący kod źródłowy i przeprowadzający proces uczenia, dystrybutorzy, audytorzy, operatorzy czy użytkownicy końcowi. Czynniki, który ostatecznie spowodowałyby szkodę, może pojawić się na różnych etapach cyklu życia systemu SI, co więcej, w pewnych układach sytuacyjnych może okazać się, że niewłaściwe zachowanie systemu SI spowodowane będzie równoczesnym zaistnieniem wielu czynników generowanych przez

²² Biała księga w sprawie sztucznej inteligencji. Europejskie podejście do doskonałości i zaufania, COM(2020) 65 final, Bruksela: Komisja Europejska 2020, s. 11. https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_pl.pdf [dostęp: 6.02.2021].

²³ Zob. J. FJELD, N. ACHTEN, H. HILLIGOSS, A. NAGY, M. SRIKUMAR, *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI*, Cambridge: Berkman Klein Center for Internet & Society 2020, s. 56-57.

różne podmioty (problem rozmycia odpowiedzialności przez udział wielu podmiotów w literaturze określany jest jako *many hands problem*²⁴).

Z jednej strony słuszne wydaje się twierdzenie, że twórca systemu powinien być odpowiedzialny za skutki spowodowane przez swoje dzieło. Przypisanie odpowiedzialności karnej za szkody wyrządzone przez system SI właśnie jemu nie budzi wątpliwości, o ile przyczyna zaistniałego skutku tkwiła w konstrukcji samego systemu i nie była spowodowana złośliwym oddziaływaniem osób trzecich. Rozwiązanie to powoduje już jednak wątpliwości w odniesieniu do skutków wywołanych przez system SI, których przyczyną było jego uczenie się już po przekazaniu przez producenta innemu podmiotowi. Nie sposób bowiem wymagać od twórcy, aby przewidział wszelkie dające się pomyśleć sytuacje, w których system może się znaleźć. Ponadto problematyczną kwestią może okazać się wykazanie związku przyczynowego pomiędzy działaniem twórcy a powstałym skutkiem. Opierając się na koncepcji obiektywnego przypisania skutku, stwierdzić należy, że przypisanie zaistniałego skutku danemu podmiotowi jest możliwe tylko wtedy, gdy miał on dostęp do odpowiednich informacji umożliwiających rozpoznanie potencjalnego skutku, albo chociaż na podstawie których skutek był możliwy do przewidzenia²⁵. Jeżeli więc w trakcie tworzenia systemu SI jego twórca obiektywnie nie mógł przewidzieć skutków, które wynikałyby z późniejszego funkcjonowania systemu, to na tej podstawie odpowiedzialności karnej ponieść nie może²⁶.

Nieco inne rozwiązanie zasugerowane zostało przez Parlament Europejski. Proponuje się bowiem, że jeżeli szkoda spowodowana działaniem systemu SI miała swoją przyczynę w wadliwym szkoleniu, testowaniu bądź konstrukcji systemu, to podmiotu odpowiedzialnego należy poszukiwać po stronie twórcy, natomiast gdy szkoda była skutkiem naruszenia zasad ostrożności przez użytkownika, to jemu należałoby przypisać odpowiedzialność²⁷. Parla-

²⁴ Zob. G. CONTISSA, *Automation and Liability: an Analysis in the Context of Socio-Technical Systems*, i-lex 2017, s. 29, http://www.i-lex.it/articles/volume11/fascicolo1/contissa_automation_and_liability.pdf [dostęp: 6.02.2021]; M. COECKELBERGH, *Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability*, „Science and Engineering Ethics” 2020, nr 26, s. 2060.

²⁵ J. GIEZEK, *Teorie związku przyczynowego oraz koncepcje obiektywnego przypisania*, [w:] *System Prawa Karnego*, t. III: *Nauka o przestępstwie. Zasady odpowiedzialności*, red. R. Dębski, Warszawa: Wydawnictwo C.H. Beck 2017, s. 547-548.

²⁶ Zob. Ł. POHL, *Przyczynek do rozważań o strukturze nieumyślności*, [w:] *Obiektywne oraz subiektywne przypisanie odpowiedzialności karnej*, red. J. Giezek, P. Kardas, Warszawa: Wolters Kluwer Polska 2016, s. 422-423.

²⁷ Rezolucja Parlamentu Europejskiego z dnia 16 lutego 2017 r. zawierająca zalecenia dla Komisji w sprawie przepisów prawa cywilnego dotyczących robotyki (2015/2103(INL)), Dz. Urz. UE

ment Europejski zauważa także, że im większą autonomicznością cechuje się system i im dłużej trwało jego szkolenie, tym większa odpowiedzialność powinna spoczywać na osobie prowadzącej takie szkolenie. Należy również wskazać, że najnowsza propozycja regulacyjna – Wniosek Komisji Europejskiej dotyczący propozycji Rozporządzenia Parlamentu Europejskiego i Rady ustanawiającego zharmonizowane przepisy dotyczące sztucznej inteligencji – projektuje rozwiązanie, zgodnie z którym odpowiedzialność za wprowadzenie do obrotu lub oddanie do użytku systemu sztucznej inteligencji powinna brać na siebie konkretna osoba fizyczna lub prawna określona jako dostawca, niezależnie od tego, czy jest ona osobą, która zaprojektowała lub opracowała ten system²⁸. Obowiązek ten odnosi się wyłącznie do systemów SI sklasyfikowanych w załączniku do wniosku jako systemy wysokiego ryzyka. Obowiązek wskazania osoby odpowiedzialnej od strony producenta mógłby w przyszłości stanowić podstawę do opracowania mechanizmów pozwalających na pociągnięcie takiej osoby również do odpowiedzialności karnej na gruncie prawa krajowego, jednak nie mogłaby ona mieć charakteru odpowiedzialności obiektywnej, niezależnej od winy.

Wydaje się, że *de lege lata* częściowym rozwiązaniem powyższego problemu może być sięgnięcie do koncepcji gwaranta nienastąpienia skutku, wyrażonej na gruncie polskiego prawa karnego w art. 2 k.k.²⁹ W omawianym typie sytuacji (*man-on-the-loop*) treścią obowiązku nadzorującego ma być monitorowanie działania systemu SI i podjęcie działania, gdyby zidentyfikowane zostało zagrożenie. Możliwe jest więc postawienie nadzorującemu system SI zarzutu, że choć sam nie wywołał zagrożenia dla dobra prawnego, to poprzez niewypełnienie ciążącego na nim obowiązku (którego realizacja przed takim skutkiem miałyby uchronić) nie przeszkodził jego zaistnieniu³⁰. Czyn taki stanowiłby więc tzw. niewłaściwe materialne przestępstwo z zaniechania³¹. Skutkiem relewantnym prawnie może być w takim przypadku zarówno naruszenie dobra prawnego, jak i narażenie go na bezpośrednie niebezpieczeństwo³².

z 2018 r., nr C 252/239. Choć twierdzenie to wyrażone zostało w akcie dotyczącym przepisów prawa cywilnego, to wydaje się wartościowe również na gruncie prawa karnego.

²⁸ Wniosek, s. 37.

²⁹ Ustawa z dnia 6 czerwca 1997 r. – Kodeks karny, tekst jedn.: Dz.U. z 2020 r., poz. 1444.

³⁰ Zob. A. ZOLL, *Komentarz do art. 2 k.k.*, [w:] *Kodeks karny. Część ogólna. Komentarz*, t. I: *Komentarz do art. 1-52 (cz. 1)*, red. W. Wróbel, A. Zoll, LEX, teza 10; D. Tokarczyk, *Obowiązek gwaranta w prawie karnym*, „Ruch Prawniczy, Ekonomiczny i Socjologiczny” 76 (2014), nr 4, s. 211.

³¹ M. KLIŚ, *Źródła obowiązku gwaranta w polskim prawie karnym*, „Czasopismo Prawa Karnego i Nauk Penalnych” III (1999), z. 2, s. 171.

³² Ł. POHL, *Komentarz do art. 2 k.k.*, [w:] *Kodeks karny. Komentarz*, red. R.A. Stefański, LEGALIS, teza 7; D. TOKARCZYK, *Obowiązek gwaranta*, s. 210.

Zgodnie z treścią art. 2 k.k. gwarantem może być tylko osoba, na której ciążył prawny, szczególnie obowiązek zapobiegnięcia skutkowi. Oznacza to, że zobowiązanie takie musi wynikać wprost z normy prawnej bądź z aktu mającego znaczenie prawne. W doktrynie nie ma jednomyślności³³ co do tego, jakie kategorie źródeł obowiązku spełniają to kryterium, choć wydaje się, że powszechnie akceptowanymi są: ustawa (czy szerzej – akt normatywny o charakterze powszechnie obowiązującym³⁴) i dobrowolne przyjęcie na siebie obowiązku zapobiegania skutkowi. Szczegółność obowiązku odnosić natomiast należy do kręgu jego adresatów. Obowiązek ma szczególny charakter wtedy, gdy obejmuje nie każdą osobę, która znalazła się w danej sytuacji, a jedynie taką, która charakteryzuje się określonymi cechami, wyróżniającymi je z uwagi na stosunek do dobra chronionego normą prawną³⁵.

Kolejnym rysującym się na tej płaszczyźnie problemem może być ustalenie obiektywnej możliwości zapobiegnięcia skutkowi. Ocena taka w kontekście funkcjonowania systemów SI obejmuje ustalenie, czy człowiek był w stanie zidentyfikować zagrożenie, nim zmaterializowało się ono w postaci narażenia dóbr prawnie chronionych, oraz czy obiektywnie był w stanie temu zapobiec. Zrozumienie działania systemów SI i przewidzenie ich decyzji stanowi wyzwanie, szczególnie uwzględniając fakt, że kod systemu może stanowić tajemnicę handlową oraz biorąc pod uwagę to, że niekiedy nie jest możliwe ustalenie, dlaczego system podejmuje określone decyzje. Ponadto systemy SI operują na wielkich zbiorach danych w tempie niekiedy niemożliwym do weryfikacji ich działań przez człowieka. Granicą obowiązku gwaranta jest więc obiektywna możliwość jego zrealizowania.

Wydaje się więc, że *de lege lata* możliwe jest przypisanie odpowiedzialności karnej nadzorującemu system SI, gdy na nadzorującym ciążył prawny szczególnie obowiązek zapobiegnięcia temu skutkowi, o ile spełnione zostaną pozostałe przesłanki przypisania odpowiedzialności karnej. Trudności natomiast powoduje przyjęcie tej konstrukcji jako podstawy odpowiedzialności użytkownika systemu SI, który korzysta z niego jako konsument. Problem ten może przybrać na znaczeniu wraz z wprowadzeniem na rynek pojazdów autonomicznych, docelowo pozbawionych kierującego. Jak trafnie wskazuje L. Bosek, systemy SI są testowane i certyfikowane przed wprowa-

³³ Zob. A. ZOLL, *Komentarz do art. 2 k.k.*, teza 17; J. LACHOWSKI, *Komentarz do art. 2 k.k.*, [w:] *Kodeks karny. Komentarz*, red. V. Konarska-Wrzošek, LEX, teza 5.

³⁴ M. KLIŚ, *Źródła obowiązku*, s. 170.

³⁵ Tamże, s. 173; A. ZOLL, *Komentarz do art. 2 k.k.*, teza 16.

dzeniem do obrotu, na nikim zaś nie ciąży powszechny obowiązek ustawicznego badania, czy dopuszczone do obiegu systemy są bezpieczne – nie można na nikogo nakładać obowiązku badania kwalifikacji zawodowców³⁶.

Jeżeli więc na użytkowniku nie ciąży prawny obowiązek związany z nadzorowaniem systemu SI, nie będzie on ponosił odpowiedzialności karnej za naruszenie dóbr prawnych przez taki system. Odnotować jednak należy, że ustawodawca wprowadza szczegółowe obowiązki w przepisach prawnych regulujących określone obszary wykorzystywania systemów SI, np. w zakresie testowania pojazdów autonomicznych³⁷ czy używania bezzałogowych statków powietrznych (BSP)³⁸. Przepisy te, jak się wydaje, mogą posłużyć do rekonstrukcji cech indywidualizujących gwaranta i treści jego obowiązków, jako przepisy modyfikujące art. 2 k.k., który jest przepisem zrębowym niezupełnym logicznie³⁹.

4. W PEŁNI AUTONOMICZNY SYSTEM SI

Trzecia grupa sytuacji, które, jak się wydaje, stanowić mogą największe wyzwanie dla prawa karnego, to przypadki, w których szkody w dobrach prawnie chronionych spowodowane zostały działaniem w pełni autonomicznych systemów SI. Choć tego typu systemy współcześnie nie są dopuszczane jeszcze do użytku, to już istnieją. Z perspektywy technicznej te systemy SI, które wymagają nadzoru człowieka, są już *de facto* autonomiczne, zaś człowiek ma pełnić rolę „bezpiecznika”, który zareaguje w odpowiedni sposób, gdy pojawi się zagrożenie⁴⁰.

Standardy projektowania i użytkowania systemów SI wypracowywane na gruncie Unii Europejskiej zmierzają ku stworzeniu mechanizmów praktycznego rozwijania koncepcji „SI godnej zaufania”. Jednym z warunków funkcjo-

³⁶ L. BOSEK, *Perspektywy rozwoju*, s. 12-13.

³⁷ Art. 65n ust. 1 pkt 2 ustawy z dnia 20 czerwca 1997 r. – Prawo o ruchu drogowym, tekst jedn.: Dz.U. z 2020 r., poz. 110.

³⁸ Załącznik nr 6b do rozporządzenia Ministra Transportu, Budownictwa i Gospodarki Morskiej z dnia 26 marca 2013 r. w sprawie wyłączenia zastosowania niektórych przepisów ustawy – Prawo lotnicze do niektórych rodzajów statków powietrznych oraz określenia warunków i wymagań dotyczących używania tych statków, tekst jedn.: Dz.U. z 2019 r., poz. 1497.

³⁹ Zob. Ł. POHL, *Artykuł 2 Kodeksu karnego w roli wyznacznika przestępstwa skutkowego wypełnionego przez zaniechanie*, „Ruch Prawniczy, Ekonomiczny i Socjologiczny” 70 (2008), nr 3, s. 92, 94.

⁴⁰ Zob. J. KACZMAREK, A. SAMPOLSKI, *Wybrane zagadnienia odpowiedzialności karnej pojazdów autonomicznych*, „Monitor Prawniczy” 2018, nr 9, s. 492.

nowania takiej SI jest nadzorcza rola człowieka. Nawet jeżeli system SI będzie z perspektywy technicznej całkowicie autonomiczny w tym sensie, że człowiek nie będzie niezbędnym elementem jego procesu decyzyjnego i realizacji decyzji, to jednak człowiek w procesie decyzyjnym musi uczestniczyć, sam system SI nie może bowiem ponosić odpowiedzialności za wywołane przez siebie skutki⁴¹. Podobne założenia przyjęto również w polskich dokumentach politycznych wyznaczających kierunki rozwoju sztucznej inteligencji. Wskazuje się w nich wprost na potrzebę supremacji człowieka nad maszyną⁴².

Najistotniejszym problemem związanym z przypisaniem odpowiedzialności za skutki spowodowane funkcjonowaniem w pełni autonomicznych systemów SI jest zerwanie łańcucha odpowiedzialności (*accountability gap*⁴³). Człowiek bowiem ani nie wydaje szczegółowych poleceń, ani nie nadzoruje działania systemu. Związek pomiędzy działaniem systemu SI a świadomością i wolą człowieka w przypadku systemów w pełni autonomicznych w sensie modelu eksploatacji (ale nie w sensie sprawstwa twórcy czy cyberprzestępcy) całkowicie zanika.

W Polsce *de lege lata* nie jest możliwe uznanie samego systemu SI za sprawcę czynu zabronionego, ten bowiem może być popełniony jedynie przez człowieka. W takich sytuacjach próbuje się niekiedy przypisać odpowiedzialność karną za wszelkie szkody temu, który taki system uruchomił⁴⁴. Pojawiają się także głosy postulujące rozważenie nadania systemom SI w przyszłości pewnej postaci podmiotowości prawnej, choć głównie formułowane są one w odniesieniu do problemów generowanych przez systemy SI na gruncie prawa cywilnego⁴⁵. Istnieją jednak stanowiska w literaturze zagranicznej, aprobujące uznanie systemu SI za podmiot zdolny do ponoszenia

⁴¹ *Wytyczne w zakresie etyki dotyczące godnej zaufania sztucznej inteligencji*, Grupa ekspertów wysokiego szczebla ds. sztucznej inteligencji, Bruksela: Komisja Europejska 2019, s. 20, https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/JURI/DV/2019/11-06/Ethics-guidelines-AI_PL.pdf [dostęp: 6.02.2021]; *Biała księga*, s. 11.

⁴² *Założenia do strategii AI w Polsce*, Warszawa: Ministerstwo Cyfryzacji 2018, s. 131; Uchwała nr 196 Rady Ministrów z dnia 28 grudnia 2020 r. w sprawie ustanowienia „Polityki dla rozwoju sztucznej inteligencji w Polsce od roku 2020”, *Monitor Polski* z 2021 r., poz. 23, załącznik, s. 66, 71.

⁴³ Zob. B.-J. KOOPS, M. HILDEBRANDT, D. JAQUETT-CHIFFELLE, *Bridging the Accountability Gap: Rights for New Entities in the Information Society?*, „Minnesota Journal of Law, Science and Technology” 11 (2010), nr 2, s. 501.

⁴⁴ Tak *de lege lata* w odniesieniu do w pełni autonomicznych pojazdów: K. MAMAK, *Prawo karne przyszłości*, Warszawa: Wolters Kluwer Polska 2017, s. 97.

⁴⁵ Zob. A. KRASUSKI, *Status prawny sztucznego agenta. Podstawy prawne zastosowania sztucznej inteligencji*, Warszawa: Wydawnictwo C.H. Beck 2021, s. 378.

odpowiedzialności karnej⁴⁶. Odwołują się one w szczególności do konstrukcji odpowiedzialności podmiotów zbiorowych za czyny zabronione, która w niektórych państwach, inaczej niż w Polsce, przybiera wprost postać odpowiedzialności karnej⁴⁷. Wprowadzenie takich rozwiązań w życie powodowałoby jednak konieczność całkowitego przemodelowania prawa karnego i jego fundamentalnych zasad. Redefinicji wymagałyby podstawowe kategorie pojęciowe, takie jak „świadomość”, „wola”, „wina”. Pojawiałyby się też pytania szczegółowe: W jaki sposób „karać” system SI i czy w ogóle kara w takim przypadku realizowałaby swoje cele?⁴⁸ Czy systemom SI w postępowaniu karnym przysługiwałyby jakieś gwarancje, jak na przykład zakaz zmuszania do samooskarżania, czy prawo do milczenia?⁴⁹ Innym rozwiązaniem mogłoby być stworzenie zupełnie nowej gałęzi prawa, regulującej stosunki prawne z udziałem systemów SI i zasady odpowiedzialności za wywołane przez nie skutki w bardziej adekwatny sposób, obejmując nie tylko sytuacje wywołane przyczynami tkwiącymi wewnątrz systemu, ale także zewnętrznym złośliwym oddziaływaniem innego podmiotu na system. Pozwoliłoby to z jednej strony na zachowanie integralności dotychczas funkcjonującego prawa karnego, z drugiej zaś – umożliwiłoby wyjście poza ciasny gorset zasad odpowiedzialności formułowanych w tej gałęzi prawa⁵⁰.

WNIOSKI

Upowszechnianie się systemów SI w życiu społecznym i ciągły rozwój tej dziedziny niosą ze sobą korzyści nie do przecenienia, ale równocześnie generują istotne problemy. Trwająca rewolucja technologiczna zmienia zasady, według których toczy się życie społeczne, tworząc wyzwania dla funkcjonujących dotychczas systemów normatywnych regulujących to życie. Skutki takie obserwowalne są również na gruncie prawa karnego. Wraz ze wzrostem

⁴⁶ G. HALLEVY, *When Robots Kill. Artificial Intelligence under Criminal Law*, Boston: Northeastern University Press 2013, s. 177-178.

⁴⁷ G. HALLEVY, *Liability for Crimes Involving Artificial Intelligence Systems*, Cham–Heidelberg–New York–Dordrecht–London: Springer 2015, s. 42-43.

⁴⁸ K. MAMAK, *Prawo karne przyszłości*, s. 89-90.

⁴⁹ Zob. A. CHŁOPECKI, *Sztuczna inteligencja – szkice prawnicze i futurologiczne*, Warszawa: Wydawnictwo C.H. Beck 2018, s. 41.

⁵⁰ R. REJMANIAK, *Odpowiedzialność karna za skutki spowodowane przez autonomiczne roboty bojowe – zarys problemu*, [w:] *Wykorzystanie dronów i robotów w systemach bezpieczeństwa. Wybrane aspekty*, red. R. Kamprowski, M. Skarżyński, Poznań: Wydawnictwo Naukowe Wydziału Nauk Politycznych i Dziennikarstwa 2019, s. 44.

poziomu autonomiczności systemów SI rosną też trudności ze wskazaniem podmiotu, któremu należałoby przypisać odpowiedzialność karną.

W przypadku systemów SI o niskim stopniu autonomiczności, z reguły odpowiedzialność będzie ciążyć na użytkowniku, do którego będzie ostatecznie należała decyzja o podjęciu działania sugerowanego przez system. Monitorowanie działalności użytkownika może odbywać się na podstawie prawa autora do sprawowania nadzoru autorskiego, które nie jest przywilejem, a obowiązkiem twórcy, pozwalającym na nadzorowanie sposobu wykorzystania i eksploatacji własnego utworu. Jeżeli natomiast rola człowieka sprowadza się jedynie do pełnienia funkcji nadzorującego system SI, to w tym przypadku przydatna wydaje się konstrukcja gwaranta, o ile na nadzorującym spoczywał prawny szczególny obowiązek zapobiegnięcia skutkowi. Zarzutem stawianym nadzorującemu jest więc nie to, że skutek wywołał, ale że mu nie zapobiegł, choć był do tego zobowiązany. Konstrukcja ta ma jednak swoje wady, nie jest bowiem przydatna w sytuacjach, w których obowiązek zapobiegnięcia skutkowi nie wynika ani z normy prawnej, ani z dobrowolnego zobowiązania, gdy użytkownikiem takiego systemu jest konsument.

Problemy związane z przypisaniem odpowiedzialności karnej za skutki spowodowane przez w pełni autonomiczne systemy SI wydają się najtrudniejsze ze względu na zerwanie związku między wolą i świadomością człowieka a działaniem systemu, skutkującym powstaniem tzw. luki w odpowiedzialności (*accountability gap*). Choć kwestie te zdają się mniej palące od określenia zasad odpowiedzialności karnej w pozostałych omówionych sytuacjach, nie znaczy to jednak, że nie są one warte rozważania. Problemy te stanowią bowiem nowe wyzwania dla prawa i tworzą nową płaszczyznę dyskusji o perspektywach i kierunkach ewolucji pojmowania podmiotowości prawnej⁵¹.

PIŚMIENNICTWO

Źródła prawa

Rozporządzenie Ministra Transportu, Budownictwa i Gospodarki Morskiej z dnia 26 marca 2013 r. w sprawie wyłączenia zastosowania niektórych przepisów ustawy – Prawo lotnicze do niektórych rodzajów statków powietrznych oraz określenia warunków i wymagań dotyczących używania tych statków, tekst jedn.: Dz.U. z 2019 r., poz. 1497.

Uchwała nr 196 Rady Ministrów z dnia 28 grudnia 2020 r. w sprawie ustanowienia „Polityki dla rozwoju sztucznej inteligencji w Polsce od roku 2020”, Monitor Polski z 2021 r., poz. 23.

⁵¹ T. PIETRZYKOWSKI, *Ludzkie, niezbyt ludzkie. Esej o podmiotowości prawnej i wyzwaniach XXI w.*, Katowice: Wydawnictwo Uniwersytetu Śląskiego 2016, s. 141.

Ustawa z dnia 20 czerwca 1997 r. – Prawo o ruchu drogowym, tekst jedn.: Dz.U. z 2020 r., poz. 110.

Ustawa z dnia 6 czerwca 1997 r. – Kodeks karny, tekst jedn.: Dz.U. z 2020 r., poz. 1444.

Literatura

- Biała księga w sprawie sztucznej inteligencji. Europejskie podejście do doskonałości i zaufania*, COM(2020) 65 final, Bruksela: Komisja Europejska 2020, https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_pl.pdf [dostęp: 6.02.2021].
- BODEN Margaret Ann: *Sztuczna inteligencja. Jej natura i przyszłość*, Łódź: Wydawnictwo Uniwersytetu Łódzkiego 2020.
- BOSEK Leszek: *Perspektywy rozwoju odpowiedzialności cywilnej za inteligentne roboty*, „Forum Prawnicze” 52 (2019), nr 2, s. 3-17.
- CHŁOPECKI Aleksander: *Sztuczna inteligencja – szkice prawnicze i futurologiczne*, Warszawa: Wydawnictwo C.H. Beck 2018.
- COECKELBERGH Mark: *Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability*, „Science and Engineering Ethics” 2020, nr 26 s. 2051-2068.
- CONTISSA Giuseppe: *Automation and Liability: an Analysis in the Context of Socio-Technical Systems*, i-lex 2017.
- CUMMINGS Mary L.: *Automation and Accountability in Decision Support System Interface Design*, „The Journal of Technology Studies” 32 (2006), nr 1, s. 23-31.
- DAVENPORT Thomas, KALAKOTA Ravi: *The potential for artificial intelligence in healthcare*, „Future Healthcare Journal” 6 (2019), nr 2, s. 94-98.
- Definicja SI: Główne funkcje i dyscypliny*, Grupa ekspertów wysokiego szczebla ds. sztucznej inteligencji, Bruksela: Komisja Europejska 2019, https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/JURI/DV/2019/11-06/Ethics-guidelines-AI_PL.pdf [dostęp: 6.02.2021].
- FILIPKOWSKI Wojciech: *Prawo karne wobec sztucznej inteligencji*, [w:] *Prawo sztucznej inteligencji*, red. L. Lai, M. Świerczyński, Warszawa: Wydawnictwo C.H. Beck 2020, s. 113-128.
- FJELD Jessica, ACHTEN Nele, HILLIGOSS Hannah, NAGY Adam, SRIKUMAR Madhu: *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI*, Cambridge: Berkman Klein Center for Internet & Society 2020.
- FLASIŃSKI Mariusz: *Wstęp do sztucznej inteligencji*, Warszawa: Wydawnictwo Naukowe PWN 2020.
- GIEZEK Jacek: *Teorie związku przyczynowego oraz koncepcje obiektywnego przypisania*, [w:] *System Prawa Karnego*, t. III: *Nauka o przestępstwie. Zasady odpowiedzialności*, red. R. Dębski, Warszawa: Wydawnictwo C.H. Beck 2017, s. 444-553.
- HALLEVY Gabriel: *Liability for Crimes Involving Artificial Intelligence Systems*, Cham–Heidelberg–New York–Dordrecht–London: Springer 2015.
- HALLEVY Gabriel: *When Robots Kill. Artificial Intelligence under Criminal Law*, Boston: Northeastern University Press 2013.
- KACZMAREK Janusz, SAMPOLSKI Adam: *Wybrane zagadnienia odpowiedzialności karnej pojazdów autonomicznych*, „Monitor Prawniczy” 2018, nr 9, s. 490-496.
- KASPERSKA Anna: *Problemy zastosowania sztucznych sieci neuronalnych w praktyce prawniczej*, „Przegląd Prawa Publicznego” 2017, nr 11, s. 25-32.
- KLIŚ Maciej: *Źródła obowiązku gwaranta w polskim prawie karnym*, „Czasopismo Prawa Karnego i Nauk Penalnych” III (1999), z. 2, s. 169-196.

- KOOPS Bert-Jaap, HILDEBRANDT Mireille, JAQUETT-CHIFELLE David-Olivier: *Bridging the Accountability Gap: Rights for New Entities in the Information Society?*, „Minnesota Journal of Law, Science and Technology” 11 (2010), nr 2, s. 497-561.
- KRASUSKI Andrzej: *Status prawny sztucznego agenta. Podstawy prawne zastosowania sztucznej inteligencji*, Warszawa: Wydawnictwo C.H. Beck 2021.
- LACHOWSKI Jerzy: *Komentarz do art. 2 k.k.*, [w:] *Kodeks karny. Komentarz*, red. V. Konarska-Wrzosek, LEX.
- LELE Ajey: *Debating Lethal Autonomous Weapons Systems*, „Journal of Defense Studies” 13 (2019), nr 1, s. 51-70.
- MAMAK Kamil: *Prawo karne przyszłości*, Warszawa: Wolters Kluwer Polska 2017.
- MASSEY Gary, EHRENSBERGER-DOW Maureen: *Machine learning: Implications for translator education*, „Lebende Sprachen” 62 (2017), nr 2, s. 300-312.
- MCCARTHY John, MINSKY Marvin Lee, ROCHESTER Nathaniel, SHANNON Claude Elwood: *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, 1955, <http://www-formal.stanford.edu/jmc/history/dartmouth.pdf> [dostęp: 6.02.2021].
- MICHIE Donald: *Methodologies from Machine Learning in Data Analysis and Software*, „The Computer Journal” 34 (1991), nr 6, s. 559-565.
- NEFF Gina, NAGY Peter: *Talking to Bots: Symbiotic Agency and the Case of Tay*, „International Journal of Communication” 2016, nr 10, s. 4915-4931.
- NEUMANN Tomasz, *Perspektywy wykorzystania pojazdów autonomicznych w transporcie drogowym w Polsce*, „Autobusy” 2018, nr 12, s. 787-794.
- PARASURAMAN Raja, SHERIDAN Thomas B., WICKENS Christopher D.: *A Model for Types and Levels of Human Interaction with Automation*, „IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans” 30 (2000), nr 3, s. 286-297.
- PIETRZYKOWSKI Tomasz: *Ludzkie, niezbyt ludzkie. Esej o podmiotowości prawnej i wyzwaniach XXI w.*, Katowice: Wydawnictwo Uniwersytetu Śląskiego 2016.
- POHL Łukasz: *Artykuł 2 Kodeksu karnego w roli wyznacznika przestępstwa skutkowego popełnionego przez zaniechanie*, „Ruch Prawniczy, Ekonomiczny i Socjologiczny” 70 (2008), nr 3, s. 87-95.
- POHL Łukasz: *Komentarz do art. 2 k.k.*, [w:] *Kodeks karny. Komentarz*, red. R.A. Stefański, LEGALIS.
- POHL Łukasz: *Przyczynek do rozważań o strukturze nieumyślności*, [w:] *Obiektywne oraz subiektywne przypisanie odpowiedzialności karnej*, red. J. Giezek, P. Kardas, Warszawa: Wolters Kluwer Polska 2016, s. 418-431.
- REJMANIAK Rafał: *Odpowiedzialność karna za skutki spowodowane przez autonomiczne roboty bojowe – zarys problemu*, [w:] *Wykorzystanie dronów i robotów w systemach bezpieczeństwa. Wybrane aspekty*, red. R. Kamprowski, M. Skarżyński, Poznań: Wydawnictwo Naukowe Wydziału Nauk Politycznych i Dziennikarstwa 2019, s. 27-50.
- Rezolucja Parlamentu Europejskiego z dnia 16 lutego 2017 r. zawierająca zalecenia dla Komisji w sprawie przepisów prawa cywilnego dotyczących robotyki (2015/2103(INL)), Dz. Urz. UE z 2018 r., nr C 252/239.
- RIBEIRO Marco Tulio, SINGH Sameer, GUESTRIN Carlos: *„Why Should I Trust You?” Explaining the Predictions of Any Classifier*, San Francisco: The 22nd ACM SIGKDD International Conference 2016, s. 1-10, <https://www.kdd.org/kdd2016/papers/files/rfp0573-ribeiroA.pdf> [dostęp: 6.02.2021].

- RÓŻANOWSKI Krzysztof: *Sztuczna inteligencja: rozwój, szanse i zagrożenia*, „Zeszyty Naukowe Warszawskiej Wyższej Szkoły Informatyki” 2007, nr 2, s. 109-135.
- RUTKOWSKI Leszek: *Metody i techniki sztucznej inteligencji*, Warszawa: Wydawnictwo Naukowe PWN 2012.
- TEGMARK Max: *Życie 3.0. Człowiek w erze sztucznej inteligencji*, Warszawa: Prószyński i S-ka 2019.
- TOKARCZYK Damian: *Obowiązek gwaranta w prawie karnym*, „Ruch Prawniczy, Ekonomiczny i Socjologiczny” 76 (2014), nr 4, s. 203-214.
- TURING Alan Mathison: *Computing Machinery and Intelligence*, „Mind” 59 (1950), nr 236, s. 433-460.
- VINGE Vernor: *The Coming Technological Singularity: How to Survive in the Post-Human Era*, [w:] *Vision-21: Interdisciplinary Science and Engineering in the Era of Cyberspace*, red. G. Landis, S. Bailey, Ohio: NASA Conference Publication 10129, NASA Lewis Research Center 1993 s. 11-22.
- Wniosek – Rozporządzenie Parlamentu Europejskiego i Rady ustanawiające zharmonizowane przepisy dotyczące sztucznej inteligencji (Akt w sprawie sztucznej inteligencji) i zmieniające niektóre inne akty ustawodawcze Unii. {SEC(2021) 167 final} – {SWD(2021) 84 final} – {SWD(2021) 85 final} COM(2021) 206 final, 2021/0106 (COD)*, Bruksela: Komisja Europejska 2021, https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0012.02/DOC_1&format=PDF [dostęp: 13.07.2021].
- Wytyczne w zakresie etyki dotyczące godnej zaufania sztucznej inteligencji*, Grupa ekspertów wysokiego szczebla ds. sztucznej inteligencji, Bruksela: Komisja Europejska 2019, https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/JURI/DV/2019/11-06/Ethics-guidelines-AI_PL.pdf [dostęp: 6.02.2021].
- Założenia do strategii AI w Polsce*, Warszawa: Ministerstwo Cyfryzacji 2018.
- ZOLL Andrzej: *Komentarz do art. 2 k.k.*, [w:] *Kodeks karny. Część ogólna. Komentarz*, t. I: *Komentarz do art. 1-52 (cz. 1)*, red. W. Wróbel, A. Zoll, LEX.

AUTONOMICZNOŚĆ SYSTEMÓW SZTUCZNEJ INTELIGENCJI JAKO WYZWANIE DLA PRAWA KARNEGO

Streszczenie

Opracowanie jest próbą przedstawienia wyzwań niesionych przez rozwój systemów sztucznej inteligencji dla prawa karnego. Wraz ze wzrostem autonomiczności systemu SI maleje rola człowieka, co generuje problem w zakresie ustalenia podmiotu, któremu należałoby przypisać odpowiedzialność karną za ewentualne narażanie lub naruszanie przez działanie systemu dóbr prawnie chronionych. W przypadku systemów o stosunkowo niskim stopniu autonomiczności, z reguły człowiek będzie odpowiedzialny za wywołane przez nie skutki. W odniesieniu do systemów bardziej zaawansowanych człowiek może ponosić odpowiedzialność jako gwarant nienastąpienia skutku, jeżeli obowiązek zapobiegnięcia skutkowi miał szczególny prawny charakter. Największe i, jak dotąd, nierozwiązane problemy z przypisaniem odpowiedzialności karnej powodować jednak będą potencjalne w pełni autonomiczne systemy SI.

Słowa kluczowe: sztuczna inteligencja; gwarant; autonomia; system SI