

Wybrane zagadnienia z zakresu usankcjonowania odpowiedzialności karnej autonomicznych robotów. Postulaty *de lege ferenda*

Słowa kluczowe: *smart robots, sztuczna inteligencja (AI), prawo karne, odpowiedzialność karna, przestępstwo, wina*

1. Wstęp

Problematyka odpowiedzialności sztucznej inteligencji za czyny zabronione pod groźbą kary jest jedną z najistotniejszych kwestii dla doktryny prawa karnego w XXI w. Technologicznie zaawansowane *smart robots*, z uwagi na możliwość podejmowania niezależnych interakcji z osobami trzecimi oraz swą zdolnością do dostosowywania i uczenia się, są po prostu nieprzewidywalne. Mogą one zatem podejmować decyzje, których efektem będzie działanie naruszające normy prawne, człowiek zaś może być albo pozbawiony możliwości przewidzenia tego zachowania, albo też nie mieć fizycznych lub intelektualnych zdolności, by do tego nie dopuścić. Odpowiedzią na powyższy stan rzeczy winno zatem być albo prze-modelowanie i zredefiniowanie podstawowych zasad i pojęć z zakresu prawa karnego, albo stworzenie całkowicie nowych i odrębnych zasad odpowiedzialności penalnej za działania autonomicznych robotów.

2. Upodmiotowienie prawne podmiotów pozaludzkich

Z uwagi na obowiązującą w Polsce zasadę *nullum crimen sine lege* uregulowanie problematyki odpowiedzialności autonomicznych robotów musiałoby nastąpić w akcie rangi ustawowej, a nie podustawowej. Mogłoby to przyjąć jedną z trzech alternatywnych form. Pierwsza polegałaby na wprowadzeniu odpowiednich regulacji do części ogólnej

ustawy – Kodeks karny¹. Druga na stworzeniu odrębnej części w ramach Kodeksu karnego, wzorem obecnej części wojskowej i sformułowaniu w niej odmiennych zasad odpowiedzialności oraz odrębnych kar. Trzecia zaś na uchwaleniu osobnej ustawy, mniej lub bardziej niezależnej od Kodeksu karnego, w której analogicznie do ustawy – Kodeks karny skarbowy² byłyby zawarte swoiste części: ogólna, szczególna i procesowa.

W opinii autorki jako najbardziej użyteczne jawi się rozwiązanie trzecie, czyli uregulowanie odpowiedzialności penalnej autonomicznych robotów samodzielnie, a więc w oderwaniu od regulacji zawartych obecnie w Kodeksie karnym. Zasadne byłoby też rozważenie, aby w ramach nowej ustawy wyraźnie wyłączyć zastosowanie przepisów części ogólnej Kodeksu karnego, do czego zobowiązuje art. 116 k.k. W dalszej kolejności należałoby określić zakres podmiotowy rzeczonych przepisów, czyli sformułować, jakie podmioty podlegają temu typu odpowiedzialności.

3. Definicja legalna autonomicznego robota

Na gruncie poczynionych wyżej ustaleń oraz mając na uwadze gwałtowny rozwój osiągnięć w dziedzinie robotyki, w następstwie którego roboty mogą być nie tylko narzędziem w rękach sprawców popełniających czyny zabronione, ale także sprawcami tych czynów, w pierwszej kolejności należałoby sformułować definicję legalną autonomicznego robota (*smart robot*)³. Potoczne rozumienie tego terminu jest bowiem bardzo szerokie i nie pozwala na wyodrębnienie różnic między wyposażonymi w sztuczną inteligencję, technicznie zaawansowanymi programami a zwykłymi, wykorzystywanymi do prostych powtarzalnych czynności maszynami. Te pierwsze to autonomiczne (*the capability for autonomy*) systemy, które dzięki algorytmom sztucznej inteligencji oraz możliwości uzyskiwania dostępu do ogromnej liczby danych są w stanie rozwijać się i uczyć (*the potential for self-learning from experience and interaction*) oraz dostosowywać swoje zachowanie do okoliczności (*the ability to adapt behaviours and actions to the environment*)⁴. Te drugie zaś to proste maszyny pozbawione możliwości wchodzenia w interakcję z ludźmi oraz niezdolne do samoczynnej zmiany swego zachowania (np. samojezdne odkurzacze, inteligentne pralki czy suszarki).

W świetle powyższych spostrzeżeń jasne jest, że wyinterpretowanie znaczenia terminu inteligentny lub autonomiczny robot z języka potocznego przy pomocy wykładni językowej nie będzie możliwe. Po pierwsze dlatego, że leksykalne jego znaczenie jest wyjątkowo niedookreślone, a w myśl zasady *nullum crimen sine lege certa* to na ustawodawcy karnym spoczywa obowiązek dostatecznej określoności czynu zabronionego. Po drugie zaś, z powodu tego, że znaczenie rzeczzonego pojęcia na potrzeby prawa karnego nie może się wyczerpywać

¹ Ustawa z 6.06.1997 r. – Kodeks karny (Dz.U. z 2025 r. poz. 383), dalej: k.k.

² Ustawa z 10.09.1999 r. – Kodeks karny skarbowy (Dz.U. z 2025 r. poz. 633).

³ Zob. § 59 pkt f Rezolucji Parlamentu Europejskiego z 16.02.2017 r. zawierającej zalecenia dla Komisji w sprawie przepisów prawa cywilnego dotyczących robotyki (2015/2103(INL)) (Dz.Urz. UE C 252, s. 239).

⁴ T. Mahler, *Smart Robotics in the EU Legal Framework. The Role of the Machinery Regulation*, „Oslo Law Review” 2024/11(1), s. 1.

tylko na płaszczyźnie semantycznej, byłaby ona bowiem zbyt uboga i nie oddawała funkcji, jaką ów autonomiczny robot może realizować przy popełnieniu czynu zabronionego.

Konkludując tę część rozważań, należy stwierdzić, że definicja normatywna robota musi uwzględniać jego autonomię oraz przyznawać mu status zbliżony do osoby prawnej, tj. uznawać go za podmiot prawa, który nie jest człowiekiem. Konieczne jest także uwzględnienie wspomnianej wyżej dychotomii, a więc faktu, że wyłącznie *smart robots* mają zdolność do samostanowienia i podejmowania autonomicznych decyzji, potencjał uczenia się, wchodzenia w interakcje z ludźmi oraz zdolność do adaptacji swego zachowania. Z punktu widzenia karnomaterialnego istotne byłoby też wyraźne podkreślenie ich niebiologicznego (*unbiological*) charakteru oraz mnogość i różnorodność funkcji (w formie przykładowego wyliczenia), które roboty te mogą pełnić.

4. Definicja przestępstwa i czynu zabronionego

W dalszej kolejności niezbędna byłaby zmiana definicji przestępstwa. Tradycyjnie rozumiane prawo karne, skoncentrowane na odpowiedzialności osoby fizycznej, nie jest bowiem w stanie odpowiednio zareagować na ewentualne zachowania bezprawne podjęte przez roboty. Z tego też powodu wszelkie tego typu działania i zaniechania są *de lege lata* bezkarne⁵.

U źródeł tego przekonania leży przede wszystkim fakt, że obecnie podmiotem przestępstwa i przedmiotem kary może być wyłącznie osoba fizyczna. To z kolei oznacza, że każda próba pociągnięcia autonomicznego robota do odpowiedzialności karnej na podstawie obecnie obowiązujących przepisów stanowiłaby przejaw stosowania analogii i wykładni rozszerzającej na niekorzyść „sprawcy”, które – jak wiadomo – są niedopuszczalne na gruncie prawa karnego (*nullum crimen sine lege stricta*). Co istotne i znamienne, przyjęte w prawie karnym pojęcie czynu także nie znajduje zastosowania do autonomicznych robotów. Karnodogmatyczna analiza treści tego terminu jednoznacznie skłania bowiem do supozycji, że może się go dopuścić wyłącznie człowiek i, co więcej, uczynić to w sposób świadomy⁶. Robot zaś *ex definitione* nie jest człowiekiem.

Na tle istniejącej regulacji niezwykle trudne, jeśli nie niemożliwe, byłoby też ustalenie stanu jego „świadomości”. Bardzo skomplikowane byłoby też udowodnienie, że świadomość

⁵ W polskiej literaturze przedmiotu zgodnie podkreśla się, że choć w art. 115 k.k. brak jest definicji legalnej przestępstwa, to jednak można ją wyinterpretować z art. 1 § 1–3 k.k. Na tej też podstawie przyjmuje się, że przestępstwo to czyn człowieka zabroniony przez ustawę pod groźbą kary, obowiązującą w czasie jego popełnienia jako zbrodnia lub występki – bezprawny, zawiniony, społecznie szkodliwy w stopniu wyższym niż znikomy. Por. L. Gardocki, *Prawo karne*, Warszawa 2021, s. 50.

⁶ W opinii Władysława Mąciora przez pojęcie czynu należy rozumieć „takie ludzkie zachowanie się, które powstaje pod wpływem aktywnej roli świadomości i jako jej produkt przeszło przez stadium decyzji podjętej ze względu na określony cel, a następnie uzewnętrzniło się w postaci działania” – zob. W. Mąciór, *Czyn ludzki i jego znaczenie w prawie karnym. Zagadnienia podstawowe*, Warszawa 1990, s. 62.

ta znalazła odzwierciedlenie w decyzji robota, a tym samym że podjęte przezeń działania lub zaniechania mogą być uznane za własne, a nie np. wymuszone przez człowieka.

Niewątpliwiej zmiany wymagałaby też sformułowana obecnie na gruncie art. 115 § 1 k.k. definicja czynu zabronionego przede wszystkim w obszarze takich przesłanek, jak postać zamiaru czy motywacja sprawcy. Analogiczne bowiem do ustalania, czy doszło do popełnienia czynu, rozpoznanie prawdziwych intencji i zamierzeń sprawczych robota w przypadku przestępstw kierunkowych, a więc tych, w których ustawodawca podkreśla cel działania sprawcy (np. oszustwo, kradzież, kradzież rozbójnicza, wymuszenie rozbójnicze etc.), byłoby niezwykle trudne⁷. Argumentem przemawiającym za takim przekonaniem jest to, że oskarżenie musiałoby nie tylko dowieść, że elementy przedmiotowe tych przestępstw mieściły się w świadomości robota, ale też iż były one objęte jego wolą.

5. Wina

Zupełnie oddzielnym zagadnieniem w aspekcie dyskusji o usankcjonowaniu odpowiedzialności penalnej autonomicznych robotów jest zakres podmiotowy przepisów o odpowiedzialności karnej. Przechodząc do szczegółów, należy w pierwszej kolejności wskazać, że obecnie wypracowane pojęcie winy nie daje się odnieść do autonomicznych robotów. Po pierwsze dlatego, że odpowiedzialność karna ukształtowana została jako odpowiedzialność za czyn własny na zasadzie winy. Tymczasem podejmowane przez roboty decyzje mają charakter zautomatyzowany i zapadają na podstawie określonego algorytmu, a nie w sposób refleksyjny. To z kolei pozwala przyjąć, że – w odróżnieniu od człowieka – robot nie ocenia ani nie „waży” poszczególnych okoliczności *in concreto*, tylko działa w taki sposób, w jaki został zaprogramowany⁸. Z tego też powodu nie potrafi on niuansować istniejących nakazów i zakazów ani właściwie rozstrzygać, który czyn charakteryzuje się wyższym stopniem społecznej szkodliwości lub też jakie dobro chronione prawem należy w danych okolicznościach poświęcić⁹. Konsekwencją powyższego jest zatem brak strony podmiotowej czynu zabronionego, a zarazem personalnej zarzucalności.

Swoistym remedium na powyższe problemy może być rekomendowane niekiedy w literaturze przedmiotu wskazanie robotowi konkretnej kwantyfikacji stopnia społecznego

⁷ S. Frankowski, *Przestępstwa kierunkowe w teorii i praktyce*, Warszawa 1970, s. 33–35.

⁸ J. Lachowski, *Prawnikarne oceny eksploatacji prostytucji uprawianej przez roboty humanoidalne*, „Ius Novum” 2024/18, s. 9.

⁹ Akademickim przykładem obrazującym ten problem jest zakodowanie robotowi ze wszech miar uzasadnionego i racjonalnego *nomen omen* zakazu „nie przejeżdżaj na czerwonym świetle”. W praktyce jednak zakaz ten może pozostawać w sprzeczności z poleceniem sformułowanym jako: „nie pozwól kierowcy zginąć w wyniku uderzenia przez nadjeżdżającą ciężarówkę”. W odróżnieniu bowiem od robotów, ludzie zdają sobie na ogół sprawę, że „nie przejeżdżaj na czerwonym świetle” nie oznacza „nigdy nie przejeżdżaj na czerwonym świetle”, lecz „nie przejeżdżaj na czerwonym świetle, chyba że masz wystarczająco dobry powód, by to uczynić”. Przytoczono za: M.A. Lemley, B. Casey, *How to Punish a Robot?*, „Stanford Law School”, <https://law.stanford.edu/wp-content/uploads/2022/07/How-to-Punish-a-Robot-CH-Abridged-MAL-BC.pdf> (dostęp: 2.12.2025 r.).

niebezpieczeństwa czynu lub wyraźne zaprezentowanie mu, jakie dobra chronione prawem zobowiązany jest strzec w pierwszej kolejności¹⁰. Zastosowanie takiego rozwiązania sprawiłoby, że robot nie musiałby samodzielnie stopniować społecznej szkodliwości czynu, lecz opierałby się na gotowej operacjonalizacji wgranej mu przez programistę. Nawet jednak i w takiej sytuacji pociągnięcie go do odpowiedzialności karnej kreowanej przez art. 1 § 3 k.k. byłoby niezwykle trudne, wymagałoby bowiem udowodnienia istnienia określonej relacji psychicznej albo stosunku psychicznego między „sprawcą” a czynem.

Przy tej okazji, mając na uwadze kompleksowość projektowanej nowelizacji, należałoby się zastanowić nad dookreśleniem przesłanek odpowiedzialności osób, które uchybiły obowiązkowi nadzoru lub kontroli nad działaniem robota. Ujmując rzecz krócej i prościej, można stwierdzić, że niekiedy przyczyną popełnienia czynu zabronionego może być trudne do przewidzenia działanie robota, innym zaś razem jego określona wada projektowa lub wdrożeniowa bądź brak odpowiedniego nadzoru. W świetle powyższego niejednoznacznie jawi się jednak kwestia tego, czy zasadne byłoby odejście od przyjętego w obecnie obowiązującej ustawie rozwiązania polegającego na indywidualizacji odpowiedzialności karnej, czy też ujęcie winy w kategoriach nieznanych prawu karnemu, a wypracowanych w prawie cywilnym. W przypadku opowiedzenia się za drugim rozwiązaniem warte rozważenia byłoby usankcjonowanie takiej konwencji normatywnej, która pozwoliłaby na poszukiwanie uzasadnienia dla niej w „szeroko rozumianej bezprawności, zamiast w winie w ujęciu podmiotowym po stronie bezpośredniego sprawcy szkody”¹¹. W takim też wypadku zasadne byłoby stworzenie czegoś na kształt winy anonimowej tudzież organizacyjnej, mogącej mieć zastosowanie wówczas, gdy niemożliwe byłoby ustalenie, która z osób ponosi odpowiedzialność za przestępcze działania lub zaniechania robota¹².

Jest konstatacją tyleż istotną, co oczywistą, że powyższe rozwiązanie mogłoby naruszać kontraktyp stypizowany w art. 27 k.k. i statuujący ryzyko nowatorstwa, a tym samym niekorzystnie wpływać na poziom postępu w szeroko rozumianej technice¹³. Niewątpliwie też wprowadzenie takiego instrumentu prawnego i przypisywanie na jego podstawie sprawstwa niezindywidualizowanej osobie stanowiłoby przykład odpowiedzialności zbiorowej albo wręcz odpowiedzialności obiektywnej (*strict or absolute liability*), wykluczającej jakiegokolwiek okoliczności ekskulpacyjne¹⁴. To zaś skutkowałoby w praktyce pociągnięciem do

¹⁰ W. Barfield, Y.H. Wengh, U. Pagallo, *The Law, Policy and Regulation for Human – Robot Interaction*, Cambridge 2024, s. 28 i n.; J.M. Balkin, *The Three Laws of Robotics on the Age of Big Data*, „Ohio State Law Journal” 2017/78, s. 17.

¹¹ Zob. szerzej: A. Głowacka, *Wina osoby prawnej. Koncepcja winy anonimowej*, „Acta Erasmiana” 2016/12(25), s. 4, https://repozytorium.uni.wroc.pl/Content/78820/PDF/03_Glowacka_A_Wina_osoby_prawnej%20_koncepcja_winy_anonimowej.pdf (dostęp: 2.12.2025 r.).

¹² Zob. szerzej: M. Bieszczad, *Znaczenie koncepcji winy organizacyjnej i winy anonimowej przy dochodzeniu roszczeń cywilnoprawnych przez pacjentów*, „Palestra” 2019/6, s. 56–58; M. Safjan, *Odpowiedzialność deliktowa osób prawnych. Stan obecny i kilka uwag de lege ferenda*, „Studia Iuridica” 1994/21, s. 187.

¹³ I. Zduński, *Przesłanki ryzyka nowatorstwa*, „Prokuratura i Prawo” 2010/10, s. 47–48.

¹⁴ M. Matecki, *Przypisanie winy. Podstawy teorii ekskulpantów*, Kraków 2019, s. 107–133.

odpowiedzialności karnej osoby, której czyn nie wypełnił znamion czynu zabronionego i której nie można postawić zarzutu zawinienia¹⁵.

Rozwiązanie to miałoby jednak dużo większą doniosłość praktyczną zwłaszcza wówczas, gdy zidentyfikowanie osób fizycznych odpowiedzialnych za przestępne działanie lub zaniechanie robota byłoby bardzo trudne. Takie sytuacje mogą zaś mieć miejsce często, albowiem proces „tworzenia” robota jest zjawiskiem o charakterze ponadnarodowym i złożonym, w którym uczestniczą inżynierowie, konstruktorzy czy mechanicy, a ich czynności pozostają ze sobą w ścisłym związku. Oceny stanu faktycznego nie ułatwia także asymetria informacji, anonimowość uczestników biorących udział w procesie tworzenia oraz fakt, że podejmowane przez *smart robots* decyzje coraz częściej są konsekwencją algorytmów o ogromnej złożoności i niejasności. W rezultacie pomimo że odstąpienie od zasad odpowiedzialności osobistej i indywidualnej stanowiłoby całkowite odejście od polskich kanonów prawa karnego, to jednak pozwoliłby na normatywne zaakcentowanie, że nadzór ma kluczowe znaczenie przy posługiwaniu się sztuczną inteligencją. Drugiego rodzaju skutkiem wprowadzenia takiego rozwiązania normatywnego byłoby zapewne „wyhamowanie” rozwoju technologicznego AI; jej producenci, producenci oprogramowania czy konstruktorzy prawdopodobnie wprowadziliby bowiem jakieś limity samouczenia się, aby zapobiec niedozwolonym lub nieoczekiwanym zachowaniom tych podmiotów.

6. Sankcje penalne

Na tle przeprowadzonych powyżej rozważań zasadne wydaje się sformułowanie jeszcze jednego postulatu *de lege ferenda*. Sprowadza się on do tego, aby opracować skuteczne, proporcjonalne i odstrasżające środki zaradcze lub sankcje karne wobec AI. Specyfika odpowiedzialności, do jakiej można pociągnąć autonomicznego robota, sprawia, że katalog omawianych środków penalnym należałoby podzielić na dwie grupy.

Do pierwszej z nich należałyby te, które umożliwiałyby zrehabilitowanie pokrzywdzonemu poniesionych strat oraz przywrócenie stanu sprzed popełnienia czynu zabronionego (*vide* polskie środki kompensacyjne stypizowane w art. 46 k.k.). W ramach tej kategorii sankcji dopuszczalne byłoby zatem orzeczenie obowiązku naprawienia, w całości albo w części, wyrządzonej przestępstwem szkody bądź zadośćuczynienia za doznaną krzywdę oraz przepadku, który umożliwiłby przejęcie przedmiotów pochodzących pośrednio z czynu zabronionego lub tych, które służyły albo były przeznaczone do jego popełnienia. Na tym tle w literaturze przedmiotu pojawiają się jednak dość często opinie, że choć wyposażenie autonomicznego robota w zdolność do ponoszenia odpowiedzialności za szkodę wyrządzoną jego „własnym” działaniem jest głęboko racjonalne oraz zgodne z założeniami sprawiedliwości naprawczej (*restorative justice*), to jednak roboty nie mają przecież własnych

¹⁵ W kontekście funkcjonowania podmiotów zbiorowych zwraca na to uwagę M. Pniewska, *Problematyka charakteru prawnego odpowiedzialności podmiotów zbiorowych za czyny zabronione pod groźbą kary w prawie polskim*, „*Studia Iuridica Toruniensia*” 2010/4, s. 189.

pieniędzy ani nie są w stanie samodzielnie wypłacić odszkodowania¹⁶. Nie polemizując z tą słuszną konkluzją, pragnę jednak poczynić zastrzeżenie, że współcześnie usprawnieniem zdolności sztucznej inteligencji nie zajmują się, jak dawniej, uniwersytety czy techniczne uczelnie wyższe, lecz przede wszystkim potężne i czerpiące ogromne zyski finansowe międzynarodowe korporacje (np. Alphabet, spółka macierzysta Google'a, czy Facebook)¹⁷. Z tego też powodu być może uzasadnione byłoby, aby producenci, operatorzy, właściciele czy użytkownicy robotów liczyli się z koniecznością wynagrodzenia szkód pozostających w określonym związku z funkcjonowaniem stworzonych przez nich lub nadzorowanych robotów¹⁸.

Do kategorii drugiej zaliczyć można by tego typu sankcje, które zobowiązywałyby *smart robots* do zaangażowania się w określone zachowanie lub powstrzymania od określonego zachowania (*vide* polskie środki karne – art. 39 k.k.). Trudność związana z ich stosowaniem polegałaby jednak na tym, że roboty nie mogą realizować działań niezapisanych w ich kodzie komputerowym. Aby zatem „zresocjalizować” robota oraz „zobowiązać” go do zgodnego z prawem działania lub zaniechania, niezbędne byłoby wgranie mu stosownych instrukcji. To z kolei wymagałoby jednoznacznego sprecyzowania, czy dane zachowanie jest tym, którego należy zakazać, do którego popełnienia należy zniechęcić, które należy opodatkować czy też na które zezwolić. Należałoby też jednoznacznie określić, kto w istniejącym łańcuchu projektowania, szkolenia, posiadania i obsługi robotów powinien być odpowiedzialny za ewentualnie wyrządzoną szkodę.

W mojej opinii nie zasługuje natomiast na aprobatę postulat, by wobec robotów stosowane były środki penalne zawierające wyraźny element sankcji i moralnego osądu (*vide* polskie kary – art. 32 k.k.). Podstawową przesłanką do takiej krytyki jest fakt, iż roboty pozbawione są możliwości doświadczania refleksji, bólu i cierpienia tym samym niemożliwe ale i bezskuteczne byłoby orzeczenie wobec nich kary stanowiącej osobistą dolegliwość.

Na koniec należałoby się także zastanowić, czy samouczące się maszyny powinny odpowiadać za wszystkie przestępstwa, tj. czy winny być w tym zakresie równe osobom fizycznym, czy też katalog takich przestępstw powinien zostać jakoś ograniczony. W opinii autorki bardziej zasadne wydaje się drugie rozwiązanie, trudno sobie bowiem wyobrazić, że robot będzie odpowiadał za przestępstwo narażenia na zarażenie (art. 161 k.k.), zmuszania

¹⁶ Zob. A. Hadzi, D. Roio, *Restorative Justice in Artificial Intelligence Crimes*, „Journal for Digital Cultures”, 20.11.2019, s. 15 i n.; M. Wenzel, T.G. Okimoto, *How Acts of Forgiveness Restore a Sense of Justice: Addressing Status/Power and Value Concerns Raised by Transgressions*, „European Journal of Social Psychology” 2010/40(3), s. 401–417.

¹⁷ A. Hadzi, D. Roio, *Restorative Justice...*, s. 16.

¹⁸ Z oczywistych względów nie będzie to łatwe. Autonomiczne roboty składają się z wielu komponentów, uczą się na podstawie bieżących interakcji z tysiącami, milionami, a nawet miliardami danych, często są projektowane, obsługiwane lub dzierżawione przez wiele różnych podmiotów, zaś wpływ na ich postawę wywierają nie tylko projektanci, lecz także osoby trzecie. Z tego też powodu ich „zachowanie” może być wyjątkowo nieprzewidywalne. Za przykład może tu posłużyć stan faktyczny, w którym chatbot Microsoftu, Tay, zaczął wygłaszać nazistowskie i homofobiczne uwagi pod adresem internautów po interakcji z tzw. trollami na ówczesnym Twitterze. Zob. szerzej: J. Vincent, *Twitter Taught Microsoft's AI Chatbot to Be a Racist Asshole in Less Than a Day*, „The Verge”, 24.03.2016, <https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist> (dostęp: 2.12.2025 r.).

do małżeństwa (art. 191 b k.k.), seksualnego wykorzystania bezradności (art. 198 k.k.), bigamii (art. 206 k.k.) czy rozpijania małoletniego (art. 208 k.k.); niewykluczone jednak, że wyobraźnia piszącej te słowa nie jest wystarczająco rozwinięta.

7. Podsumowanie

Uregulowanie odpowiedzialności karnej autonomicznych robotów za czyny zabronione pod groźbą kary budzi wiele wątpliwości natury prawnej, moralnej i filozoficznej. Wśród przedstawicieli doktryny pojawiają się niekiedy głosy, że rozważania nad takim kierunkiem zmian legislacyjnych stanowi przykład nieuzasadnionej futurologii, nie każda bowiem nowa tendencja technologiczna widoczna w rzeczywistości społecznej wymaga ingerencji karno-prawnej. Zastrzeżenia budzi też fakt, że upodmiotowienie prawne robotów jest sprzeczne z humanistyczną aksjologią otaczającego nas świata.

W opinii autorki twierdzenia te nie zasługują jednak na aprobatę. Zasygnalizowane powyżej jedynie skrótowo zagadnienia stanowią o potrzebie wprowadzenia do ustawodawstwa karnego odpowiednich regulacji umożliwiających penalizowanie bezprawnych zachowań autonomicznych robotów. Wyznacznikiem racjonalnej polityki karnej zawsze jest bowiem nadążające za aktualnymi wyzwaniami prawo karne. Tak też powinno być w tym wypadku. Wpływ nauki i nowych technologii na codzienne życie ludzi oraz rozwój przestępczości jest przecież niezaprzeczalny. By zatem prawo nie stało się anachronicznym przeżytkiem, konieczne jest wprowadzenie zmiany jego ram oraz dodania nowej kategorii podmiotów, nawet jeśli na razie jawi się to jako nieco przedwcześnie.

Abstract

Outline of the problem in the regulation of criminal liability of smart robots. Postulates de lege ferenda

Dr Izabela Jankowska-Prochot – Assistant Professor at WSB Merito University in Warsaw; expert in legislation at the National Police Headquarters and Vice President of the Management Board of the Association for Research on the Sources and Functions of Law "Fontes"
ORCID: 0000-0001-6140-0591

Technological advances and the growing autonomy of Artificial Intelligence (AI) make it necessary to enact legislation that stipulates criminal or repressive liability because the current criminal legal regulations do not allow to combat crimes related to the activities of these entities. This article examines the Polish regulations and provides analysis of legal matters connected with criminal liability for smart robots.

Keywords: smart robots, Artificial Intelligence (AI), criminal law, criminal liability, an offence, fault

Bibliografia / References

- Balkin J.M., *The Three Laws of Robotics on the Age of Big Data*, „Ohio State Law Journal” 2017/78
- Barfield W., Wengh Y.H., Pagallo U., *The Law, Policy and Regulation for Human – Robot Interaction*, Cambridge 2024
- Bieszczad M., *Znaczenie koncepcji winy organizacyjnej i winy anonimowej przy dochodzeniu roszczeń cywilnoprawnych przez pacjentów*, „Palestra” 2019/6
- Frankowski S., *Przestępstwa kierunkowe w teorii i praktyce*, Warszawa 1970
- Gardocki L., *Prawo karne*, Warszawa 2021
- Głowacka A., *Wina osoby prawnej. Koncepcja winy anonimowej*, „Acta Erasmiana” 2016/12(25), https://repozytorium.uni.wroc.pl/Content/78820/PDF/03_Glowacka_A_Wina_osoby_prawnej%20_koncepcja_winy_anonimowej.pdf
- Hadzi A., Roio D., *Restorative Justice in Artificial Intelligence Crimes*, „Journal for Digital Cultures”, 20.11.2019
- Lachowski I., *Prawnokarne oceny eksploatacji prostytucji uprawianej przez roboty humanoidalne*, „Ius Novum” 2024/18
- Lemley M.A., Casey B., *How to Punish a Robot?*, „Stanford Law School”, <https://law.stanford.edu/wp-content/uploads/2022/07/How-to-Punish-a-Robot-CH-Abridged-MAL-BC.pdf>
- Mahler T., *Smart Robotics in the EU Legal Framework. The Role of the Machinery Regulation*, „Oslo Law Review” 2024/11(1)
- Matecki M., *Przypisanie winy. Podstawy teorii ekskulpantów*, Kraków 2019
- Mąciór W., *Czyn ludzki i jego znaczenie w prawie karnym. Zagadnienia podstawowe*, Warszawa 1990
- Pniewska M., *Problematyka charakteru prawnego odpowiedzialności podmiotów zbiorowych za czyny zabronione pod groźbą kary w prawie polskim*, „Studia Iuridica Toruniensia” 2010/4
- Safjan M., *Odpowiedzialność deliktowa osób prawnych. Stan obecny i kilka uwag de lege ferenda*, „Studia Iuridica” 1994/21
- Vincent J., *Twitter Taught Microsoft's AI Chatbot to Be a Racist Asshole in Less Than a Day*, „The Verge”, 24.03.2016, <https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist>
- Wenzel M., Okimoto T.G., *How Acts of Forgiveness Restore a Sense of Justice: Addressing Status/Power and Value Concerns Raised by Transgressions*, „European Journal of Social Psychology” 2010/40(3)
- Zduński I., *Przesłanki ryzyka nowatorstwa*, „Prokuratura i Prawo” 2010/10