



Paweł Księżak

Doktor habilitowany nauk prawnych, profesor w Katedrze Prawa Cywilnego Wydziału Prawa i Administracji Uniwersytetu Łódzkiego, sędzia Sądu Najwyższego. Zajmuje się prawem sztucznej inteligencji, robotów i transhumanizmu.

✉ pkieszak@wpia.uni.lodz.pl

<https://orcid.org/0000-0002-1224-9183>



Sylwia Wojtczak

Doktor habilitowana nauk prawnych, profesor Uniwersytetu Łódzkiego w Zakładzie Polityki Prawa WPiA UŁ. Prowadzi interdyscyplinarne badania nad aksjologią prawa, językiem prawnym i prawniczym oraz mechanizmami poznania prawniczego.

✉ swojtczak@wpia.uni.lodz.pl

<https://orcid.org/0000-0001-6054-0963>

Paweł Księżak, Sylwia Wojtczak

Prawa Asimova, czyli *science fiction* jako fundament nowego prawa cywilnego

Asimov's Laws or Science Fiction as the Basis of a New Civil Law

On the 16th of February 2017 the European Parliament by the Resolution 2015/2103(INL) called on the European Commission to elaborate new solutions based on civil law that could respond to the rapid present-day development of robotics and AI. The Resolution indicates Asimov's Laws as one of the bases of civil law rules on robotics. The authors of the article analyse whether, taking the professional legal perspective, the Laws may really play such a role adequately. Their analysis reveals that because of many reasons the Laws are not good instruments for drafting legal rules. Even if the Laws are treated as very general ethical recommendations they rather cause confusions and unsolvable problems than give real directives. The authors speculate that referring to them was an expression of attachment of the EP to the old androidal paradigm of robotics assumed decades ago and still strongly present in popular culture.

Słowa kluczowe: Unia Europejska, UE, Parlament Europejski, Rezolucja 2015/2103(INL), sztuczna inteligencja, robotyka, Prawa Asimova, prawo cywilne, prawo robotów

Key words: European Union, EU, European Parliament, Resolution 2015/2103(INL), Artificial Intelligence, robotics, Asimov's Laws, civil law, robot's law

[https://doi.org/10.32082/fp.v0i4\(60\).378](https://doi.org/10.32082/fp.v0i4(60).378)

1. Wstęp

W dniu 16 lutego 2017 r. Parlament Europejski zwrócił się w formie rezolucji do Komisji z zaleceniami wypracowania rozwiązań (przepisów) w zakresie prawa

cywilnego dotyczących robotyki¹.

¹ Rezolucja Parlamentu Europejskiego z dnia 16 lutego 2017 r. zawierająca zalecenia dla Komisji w sprawie przepisów prawa cywilnego doty-

Dokument ten nadal stanowi najważniejszą wypowiedź projektującą przyszłe zmiany w zakresie prawa cywilnego na tym polu. Wśród zasad ogólnych, które rezolucja ta wymienia jako kluczowe dla przyszłego prawa cywilnego dotyczącego robotyki, wymieniono na pierwszym miejscu prawa Asimova. Fakt ten, choć oczywisty dla szerokiej publiczności, może być dla specjalistów z zakresu szeroko rozumianej sztucznej inteligencji i robotyki zaskakujący oraz stawiać pod znakiem zapytania realność czy profesjonalność oczekiwań Parlamentu Europejskiego wobec przyszłej legislacji.

Rozwój sztucznej inteligencji (ang. *artificial intelligence* – AI) i robotyki w ostatnich latach uświadomił potrzebę stworzenia ram prawnych odpowiednich dla tego nowego zjawiska. Dyskusja na ten temat jest już dość zaawansowana. Wskazuje się najważniejsze obszary, w których refleksja jest najbardziej potrzebna. Za najpilniejsze uważa się poszukiwanie etycznych fundamentów dla dalszej ekspansji AI² oraz konieczność wypracowania nowych rozwiązań odnoszących się do prawa cywilnego. Pojazdy autonomiczne, roboty medyczne czy programy doradcze stawiają fundamentalne pytania o odpowiedzialność cywilną (np. odnoszące się do winy)³; z kolei rozwój automatycz-

nych, inteligentnych systemów zawierania umów każe w nowy sposób spojrzeć na wiele podstawowych zasad i utrwalonych instytucji prawa kontraktów (takich jak np. świadomość, wola, błąd, podstęp, wykładnia umowy, dobra wiara AI). Regulacja tych kwestii szczegółowych wymaga jednak uchwycenia pewnych punktów bazowych, które staną się fundamentem dla zbudowania prawa cywilnego odnoszącego się do sztucznej inteligencji. Zrozumienie istoty problemów, które ujawniają się i mogą się ujawnić wraz z rozwojem AI, wymaga nie tylko refleksji nad znanymi mechanizmami (jak np. odpowiedzialność za produkt niebezpieczny czy odpowiedzialność na zasadzie ryzyka) i potrzebą wprowadzania pewnych korekt, lecz dyskusji odnoszącej się do rudymetów, a przez to również kierunków i celów rozwoju prawa cywilnego.

Znaczenie przywołanej rezolucji PE dla projektowania zmian w prawie cywilnym trudno przecenić. Wytycza ona ogólne ramy, w których powinni poruszać się eksperci projektujący przyszłe prawo odnoszące się do robotów na płaszczyźnie cywilnoprawnej. Każdy z elementów rezolucji zasługuje na uwagę, może bowiem wpłynąć na kształt przyszłego prawa na długie lata, a kierunki, które zostaną dziś wytyczone, bardzo trudno będzie modyfikować w przyszłości. Jest to o tyle istotne, że w rezolucji tej zwrócono się do Komisji Europejskiej, na podstawie art. 225 TFUE, o przedłożenie, na podstawie art. 114 TFUE, wniosku dotyczącego dyrektywy w sprawie przepisów prawa cywilnego odnoszących się do robotyki w oparciu o szczegółowe zalecenia zawarte w załączniku do niniejszej rezolucji. Wskazano m.in., że celem rewolucji technologicznej jest służba ludzkości (zasada przynoszenia korzyści), że ramy etyczne powinny opierać się na zasadach i wartościach zapisanych w art. 2 TUE oraz w Karcie praw podstawowych Unii Europejskiej, takich jak godność ludzka, równość, sprawiedliwość i równouprawnienie, brak dyskryminacji, świadoma zgoda, ochrona życia prywatnego i rodzinnego oraz ochrona danych, a także na innych zasadach i wartościach stanowiących podstawę prawa UE, takich jak brak stygmatyzacji, przejrzystość, autonomia, odpowiedzialność jednostki, jak również – na obowiązują-

czących robotyki, P8_TA(2017)0051 (Dz.Urz. UE C 252 z 18.7.2018, s. 239–257).

2 EC Joint Research Centre, *Artificial Intelligence. A European Perspective*, Luxembourg 2018, s. 120–121, <https://publications.jrc.ec.europa.eu/repository/bitstream/JRC113826/ai-flagship-report-online.pdf> (dostęp 22.06.2020); EC High-Level Expert Group on AI, 8.4.2019, *Ethics Guidelines for Trustworthy AI*, <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai> (dostęp 17.06.2020).

3 W tym zakresie prace są obecnie najbardziej zaawansowane, por. European Commission, Commission Staff Working Document, *Liability for emerging digital technologies* COM(2018) 237 final, Brussels, 24.4.2018, <https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX%3A52018SC0137>; Expert Group on Liability and New Technologies – New Technologies Formation, *Liability for Artificial Intelligence and Other Emerging Digital Technologies*, European Union 2019; European Commission, *Raport on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics*, Brussels, 19.2.2020, COM (2020) 64 final, [https://](https://ec.europa.eu/info/sites/info/files/report-safety-liability-artificial-intelligence-feb2020_en_1.pdf)

ec.europa.eu/info/sites/info/files/report-safety-liability-artificial-intelligence-feb2020_en_1.pdf (dostęp 29.05.2020).

cych kodeksach etycznych i praktykach stosowanych w tej dziedzinie. Podkreślono, że należy brać pod uwagę bezpieczeństwo, zdrowie i ochronę ludzi – ich wolności, prywatności, uczciwości i godności. Wskazano także niektóre środki techniczne realizacji wymienionych zasad, takie jak: rejestry robotów, fundusze odszkodowawcze, indywidualne numery rejestracyjne robotów, ubezpieczenia obowiązkowe itd.

Zakres zastosowania rezolucji ograniczono świadomością do inteligentnych robotów, tj. systemów, które posiadają następujące cechy:

- zdobywanie autonomii za pomocą czujników lub wymiany danych z otoczeniem (wzajemne połączenia) oraz wymiany i analizy tych danych;
- zdolność samokształcenia na podstawie zdobytego doświadczenia i interakcji z otoczeniem (kryterium fakultatywne);
- przynajmniej minimalna forma fizyczna;
- dostosowywanie swoich zachowania i działań do otoczenia;
- brak funkcji życiowych w sensie biologicznym.

Odnotujmy, że rozróżnienie na AI i roboty może być kwestionowane, a w dłuższej perspektywie trudne do utrzymania, ale na potrzeby niniejszej analizy jest wystarczająco precyzyjne i dopuszczalne. W dalszej części odnosić się zatem będziemy przede wszystkim do robotów, tak jak zostały one zdefiniowane powyżej.

Jak już wspomniano, rezolucja PE na pierwszym miejscu spośród zasad i wartości, które winny być brane pod uwagę przy projektowaniu nowego prawa cywilnego z zakresu robotyki, wymienia prawa Asimova. Poniższa analiza dotyczy krytycznej oceny tego zalecenia. Należy zbadać, czy prawa Asimova mogą rzeczywiście stanowić podstawę budowania systemu normatywnego. Analiza w tym zakresie pozwoli postawić ewentualną hipotezę co do możliwości przełożenia wskazanego zalecenia na konkretne rozwiązania.

2. Geneza konceptualna i teoretyczna badanego problemu

Prawa Asimova to zespół reguł wymyślonych w latach 40. i uzupełnionych w latach 80. ubiegłego stulecia przez amerykańskiego pisarza *science fiction* Isaaca Asimova. Prawa Asimova początkowo traktowano bardzo poważnie. Niektórzy uważają, że swoją popularność zawdzięczają one temu, iż były remedium

na tzw. kompleks Frankensteina, czyli obecny u jednostek i grup społecznych strach przed sztucznymi tworami człowieka, będącymi rezultatem rozwijającej się technologii (zwłaszcza tymi przypominającymi pod pewnymi względami ludzi). Kompleks Frankensteina to obawa, że uwolnią się one spod kontroli swych twórców i obrócą przeciwko człowiekowi⁴. Kompleks ten leży u podłoża różnych teorii, które przepowiadają zagładę lub niewolę ludzkości z powodu rebelii sztucznej inteligencji. O micie Frankensteina wspomina się w rezolucji PE w pierwszym w kolejności punkcie „Wprowadzenia”:

„A. mając na uwadze, że od postaci Frankensteina stworzonej przez Mary Shelley po antyczny mit o Pigmalionie, poprzez historię praskiego golema i robota Karla Čapka (autora pojęcia robot) ludzkość zawsze snuła fantazje na temat możliwości stworzenia inteligentnych maszyn, najczęściej androidów o cechach ludzkich”⁵.

Należy tu podkreślić, że choć w rezolucji nie przywołuje się wprost związanych z mitem Frankensteina negatywnych emocji, to jednak należy domniemywać, że w jakimś stopniu kierowały one jej twórcami. Świadczy o tym choćby przywołanie w pierwszym w kolejności punkcie „Zasad ogólnych” rezolucji (pkt T) praw Asimova, które – choć dzisiaj krytykowane z pozycji naukowych – stały się za sprawą publikacji i dyskusji z wczesnego okresu rozwoju robotyki i sztucznej inteligencji częścią swego rodzaju „ludowej teorii robotyki i wyobrażonej przyszłości”⁶. Są też elementem kultury popularnej i odgrywają tam rolę środka łagodzenia strachu przed nowym i nieznanym. Prawa Asimowa przewidują bowiem, co następuje:

- (0) Robot nie może skrzywdzić ludzkości lub poprzez zaniechanie doprowadzić do uszczerbku ludzkości⁷.

4 I. Asimov, *Bicentennial Man*, New York 1976; S.L. Anderson, *Asimov's "Three Laws of Robotics" and Machine Metaethics*, „AI & Society” 2008, vol. 22, nr 4, s. 477–493.

5 Rezolucja PE..., pkt A.

6 R. Barbrook, *Imaginary Futures. From Thinking Machines to the Global Village*, London 2007, s. 3–12.

7 I. Asimov, *Robots and Empire*, New York 1985.

- (1) Robot nie może skrzywdzić człowieka ani przez zaniechanie dopuścić, aby człowiek doznał krzywdy.
- (2) Robot musi być posłuszny rozkazom człowieka, chyba że stoją one w sprzeczności z pierwszym prawem.
- (3) Robot musi chronić sam siebie, jeśli to nie stoi w sprzeczności z pierwszym i drugim prawem⁸.

ma prawo kierować jego zachowaniem? Co oznacza „ludzkość”? Czy robot ma chronić ludzi i ludzkość nawet wbrew ich woli? Czy jeśli człowiek chce popełnić samobójstwo albo poświęcić swe życie dla szczytnej idei lub ochrony innego człowieka, to robot powinien temu zapobiec?

Konsekwencją tych rozważań jest też pojawienie się dylematu moralnego zawartego w następującym



Robot nie może skrzywdzić człowieka ani przez zaniechanie dopuścić, aby człowiek doznał krzywdy. Robot musi być posłuszny rozkazom człowieka, chyba że stoją one w sprzeczności z pierwszym prawem.

Należy jednak koniecznie podkreślić, że obecnie wśród specjalistów z zakresu robotyki dominuje przekonanie, że prawa Asimova są obarczone wieloma wadami i nie da się ich zastosować w praktyce⁹.

Pośród różnych argumentów padają na przykład takie, że prawa te są zbyt niejednoznaczne, żeby jakkolwiek robot mógł je pojąć (zauważmy, że są sformułowane w języku potocznym). Nawet dla ludzi nie są one do końca jasne. Bo co na przykład oznacza „krzywdzić”? Czy chodzi tylko o uszczerbek na ciele, czy też na majątku? Czy może wchodzi w grę również krzywda niematerialna? Jak nauczyć robota odróżniać działanie lub zaniechanie, które z pozoru wyglądają na krzywdzące, bo wiążą się z naruszeniem ciała lub zadawaniem bólu, ale służą dobru człowieka, od działania lub zaniechania rzeczywiście krzywdzącego? Jeśli traktujemy prawa Asimova poważnie, to jak w takim razie wytłumaczyć, że wiele robotów ma zastosowania bojowe? Jaką decyzję ma podjąć robot, jeśli z każdą wiąże się – nieunikniona – różna krzywda różnych ludzi? Dlaczego robot miałby być posłuszny rozkazom każdego człowieka, a nie wyłącznie tego, który

pytaniu: jeśli jednak robot byłby w stanie zrozumieć wszystkie abstrakcyjne pojęcia, które zawarte są w prawach Asimova, i rozwiązać ww. problemy, to dlaczego miałby być traktowany jak niewolnik człowieka? Byłby wówczas przecież do człowieka bardzo podobny. Ze strony ludzi byłby to więc przejaw rasizmu czy uprzedzenia gatunkowego¹⁰, co zasługiwałoby co najmniej na moralne potępienie, jeśli nie prawny zakaz. Z tymi wątpliwościami trzeba będzie się skonfrontować, gdy powstanie silna AI, a tym bardziej, gdy powstanie jakiś rodzaj świadomej AI.

W tym miejscu należy odnotować podstawową słabość odwołania się do praw Asimova bez jednoznacznego sprecyzowania, jakie roboty miałyby tym prawom podlegać. Wiąże się to właśnie z poziomem zaawansowania AI. Za powszechnie przyjmowany należy obecnie uznać podział w tym zakresie na co najmniej dwie podstawowe grupy, tj. wąską (albo słabą) i ogólną (albo silną) AI¹¹. Ogólna AI to system,

⁸ Tenże, *Runaround*, New York 1942.

⁹ L. McCauley, *AI Armageddon and the Three Laws of Robotics*, „Ethics and Information Technology” 2007, vol. 9, nr 2, s. 153.

¹⁰ R. Murphy, D. Woods, *Beyond Asimov: The Three Laws of Responsible Robotics*, „IEEE Intelligent Systems” 2009, vol. 24, nr 4, s. 14–20.

¹¹ J.R. Searle, *Minds, Brains and Science*, Cambridge Ma 1984, s. 26; N. Nilsson, *The Quest for Artificial Intelligence: A History of Ideas and Achievements*, New York 2009, s. 319.

który może wykonywać większość aktywności ludzkich, wąską zaś to system, który może wykonać tylko jedno lub kilka określonych zadań. Nie ulega wątpliwości, że obecnie rozwijane systemy to przykłady wąskiej AI¹²; nie jest natomiast jasne, czy i ewentualnie w jakiej perspektywie czasowej zostanie zbudowana silna AI¹³. Niekiedy wskazuje się, że różnica między słabą a silną AI nie ma charakteru jakościowego, lecz raczej ilościowy, tzn. że przejście na wyższy poziom

W odpowiedzi na te i inne pytania oraz wątpliwości podejmowane są również próby konstruowania praw konkurencyjnych. Na przykład Robin Murphy i David D. Woods zaproponowali następujące prawa robotyki:

1. Człowiek nie może wdrażać (tworzyć) robota, jeśli zastosowany system pracy w relacji człowiek – robot nie odpowiada najwyższemu prawnym i profesjonalnym standardom w zakresie bezpieczeństwa i etyki.

Czy robot ma chronić ludzi i ludzkość nawet wbrew ich woli? Czy jeśli człowiek chce poświęcić swe życie dla szczytnej idei lub ochrony innego człowieka, to robot powinien temu zapobiec?

nie nastąpi skokowo, lecz raczej w drodze ewolucji i rozwoju słabej AI¹⁴. W takim ujęciu między słabą a silną AI może istnieć wiele stopni pośrednich. Z kolei silna AI to także pojęcie niejednoznaczne. Dominuje stanowisko, że silna AI jest pierwszym krokiem do zbudowania samoświadomej superinteligencji, przy czym chwila powstania tej ostatniej (co może nastąpić w ciągu miesięcy, ale niewykluczone, że jedynie godzin od pojawienia się silnej AI) jest opisywana jako granica osobliwości, poza którą nie sposób już przewidzieć dalszego rozwoju¹⁵. Nie jest jasne, czy prawa Asimova w ujęciu rezolucji PE są związane z robotami z każdym rodzajem AI.

2. Robot musi reagować na ludzi odpowiednio do ich ról.
3. Robot musi być wyposażony w wystarczającą i odpowiednią do okoliczności autonomię, żeby mógł chronić swoją egzystencję, o ile tylko taka ochrona zapewnia sprawne przekazanie kontroli innemu podmiotom, zgodnie z pierwszym i drugim prawem¹⁶.

Jeszcze inny zestaw takich praw zaproponował Roger Clarke.

Meta-prawo: robot nie może działać, jeżeli jego działania nie są poddane prawom robotyki.

Prawo zerowe: robot nie może skrzywdzić ludzkości lub przez zaniechanie pozwolić, aby ludzkość została skrzywdzona.

Prawo pierwsze: robot nie może skrzywdzić istoty ludzkiej lub przez zaniechanie pozwolić, aby istota ludzka została skrzywdzona, chyba że to naruszałoby prawo wyższego rzędu.

Prawo drugie: robot musi wykonywać rozkazy wydane przez istotę ludzką, chyba że rozkazy te stałyby w sprzeczności z prawem wyższego rzędu. Robot musi wykonywać rozkazy wydane przez

12 EC High-Level Expert Group on AI, *A definition of Artificial Intelligence: main capabilities and scientific disciplines*, Brussels 8.4.2019, s. 5, <https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines> (dostęp 17.06. 2020).

13 N. Bostrom, *Superintelligence. Paths, Dangers, Strategies*, Oxford 2014, s. 36.

14 P. Księżak, *Zdolność prawna sztucznej inteligencji (w:) Czynić postęp w prawie. Księga jubileuszowa dedykowana Profesor Birucie Lewaszkiewicz-Petrykowskiej*, red. W. Robaczyński, Łódź 2017, s. 72.

15 N. Bostrom, *Superintelligence...*, s. 17.

16 R. Murphy, D. Woods, *Beyond Asimov...*, s. 19.

robota nadrzędnego, chyba że rozkazy te stałyby w sprzeczności z prawem wyższego rzędu.

Prawo trzecie: robot musi chronić egzystencję robota nadrzędnego, tak długo jak ta ochrona nie będzie stała w sprzeczności z prawem wyższego rzędu. Robot musi chronić swoją własną egzystencję, tak długo jak ta ochrona nie będzie stała w sprzeczności z prawem wyższego rzędu.

Prawo czwarte: robot musi wykonywać obowiązki, do których został zaprogramowany, z wyjątkiem sytuacji gdy stałoby to w sprzeczności z prawem wyższego rzędu.

Prawo prokreacji: robot nie może brać udziału w projektowaniu lub produkcji innego robota, jeżeli działania nowego robota nie będą podporządkowane prawom robotyki¹⁷.

Mimo zrelacjonowanej powyżej w skrócie dyskusji i wątpliwości rezolucja PE bezrefleksyjnie odwołuje się do praw Asimova w ich, jak się wydaje, podstawowym kształcie. Czyni to w kontekście prawa cywilnego, co – jak wskazano – sprawia dodatkowe trudności. Echo filozofii stojącej za koncepcją ogólnikowych praw Asimova odnaleźć można także w późniejszym dokumencie autorstwa grupy eksperckiej powołanej przy Komisji Europejskiej¹⁸. Autorzy dokumentu proponują wprowadzenie czterech zasad etycznych w kontekście systemów AI. Są to:

- 1) poszanowanie autonomii człowieka,
- 2) zapobieganie szkodom,
- 3) sprawiedliwość,
- 4) możliwość wyjaśnienia.

Spośród wskazanych zasad podobieństwo do praw Asimova wykazują zasada pierwsza i druga. Zostały one opisane w sposób nakazujący uznać je wyłącznie za pewne wskazówki etyczne, których przełożenie prawne na płaszczyźnie prawa cywilnego może być bardzo trudne. W odniesieniu do zasady poszanowania autonomii człowieka dokument przewiduje, co następuje:

„Prawa podstawowe, na których opiera się UE, są ukierunkowane na zapewnienie poszanowania wol-

ności i autonomii człowieka. Osoby wchodzące w interakcje z systemami AI muszą być w stanie zachować pełną i efektywną zdolność samostanowienia o sobie oraz być w stanie uczestniczyć w procesie demokratycznym. Systemy AI nie powinny bezpodstawnie podporządkowywać, przymuszać, oszukiwać, kształtować lub kontrolować ludzi ani nimi manipulować. Systemy AI powinny natomiast mieć na celu zwiększenie, uzupełnienie i wzmocnienie zdolności poznawczych, społecznych i kulturowych człowieka. Podział funkcji między ludzi i systemy AI powinien opierać się na zasadach projektowania ukierunkowanych na człowieka i pozostawiać człowiekowi rzeczywistą możliwość wyboru. Oznacza to zapewnienie sprawowania nadzoru przez człowieka i kontroli nad procesami pracy w systemach AI”¹⁹.

Z kolei zasada zapobiegania szkodom została wyjaśniona w następujący sposób:

„Systemy AI nie powinny powodować ani powiększać szkody, ani w inny sposób wywierać niekorzystnego wpływu na człowieka. Wiąże się to z ochroną godności człowieka, a także jego integralności psychicznej i fizycznej. Systemy AI oraz środowiska, w których działają, muszą być bezpieczne i pewne. Systemy te muszą być solidne pod względem technicznym, przy czym należy zapewnić, aby nie były one podatne na wykorzystanie w złym zamiarze. Należy zwrócić większą uwagę na osoby wymagające szczególnego traktowania i uwzględnić je przy opracowywaniu i wdrażaniu systemów AI. Szczególną uwagę należy zwrócić również na sytuacje, w których systemy AI mogą powodować lub pogłębiać niekorzystny wpływ spowodowany asymetrią władzy lub informacji, np. sytuacje występujące w relacjach między pracodawcami a pracownikami, przedsiębiorstwami a konsumentami, lub rządami a obywatelami. Zapobieganie szkodom wiąże się również z uwzględnianiem środowiska naturalnego i wszystkich żyjących istot”²⁰.

Obie przytoczone zasady bardziej nawet niż prawa Asimova w swej oryginalnej postaci stanowią wyraz

17 R. Clarke, *Asimov's laws of robotics: Implications for information technology. Part 2*, „Computer” 1994, vol. 27, nr 1, s. 57–66.

18 EC High-Level Expert Group on AI, 8.4.2019, *Ethics Guidelines...*, s. 12–13.

19 Tamże, s. 12.

20 Tamże.

marzenia o wszechmocnym dżinie, nad którym człowiek sprawuje pełną kontrolę. Ten fantazmat – jakże pociągający – wydaje się dostępny prawie na wyciągnięcie ręki; wystarczy jedynie doprecyzować go ustawowo. Zastanówmy się, czy jest to możliwe.

3. Analiza problemu – prawa Asimova w prawie, w szczególności prawie cywilnym

W oryginalnym ujęciu autora tych praw, pisarza I. Asimova, ich adresatem były same roboty. Trudno przypuszczać, żeby taka była również intencja Parlamentu Europejskiego. Wszak nie istnieją dzisiaj – i prawdopodobnie w najbliższym czasie nie będą istnieć

(nie użyto nawet imienia twórcy praw), prawa Asimova mają status szczególny – toposu, rozumianego albo jako miejsce wspólne (*locus communis*) dla różnych dziedzin, takich jak robotyka, etyka, prawo, albo jako argument niewymagający uzasadnienia, albo jako aksjomat będący niepodważalną podstawą jakiejś teorii.

Możliwe jest także przyjęcie stanowiska, że prawa Asimova są tu przywoływane raczej jako pewna figura retoryczna, podczas gdy tak naprawdę chodzi o nałożenie na twórców, współtwórców i kierujących robotami szczególnych obowiązków w zakresie tworzenia i wykorzystywania robotów. I choć w takim przypadku wiele z wymienionych powyżej problemów także się pojawia



Obie zasady stanowią wyraz marzenia o wszechmocnym dżinie, nad którym człowiek sprawuje pełną kontrolę. Ten fantazmat – jakże pociągający – wydaje się dostępny prawie na wyciągnięcie ręki.

(tak zresztą zakłada przywołana rezolucja PE) – roboty na tyle „rozumne”, aby uczynić je adresatami jakichkolwiek norm czy podmiotami do czegokolwiek zobowiązanymi. Obecnie zakaz krzywdzenia ludzi może być dla współczesnego robota co najwyżej wyzwaniem określonego bloku decyzyjnego zapisanego w jego oprogramowaniu. Dlatego też w rezolucji w pkt T napisano: „mając na uwadze, że prawa Asimova muszą być traktowane jako prawa adresowane do projektodawców, producentów i operatorów robotów, w tym robotów z wbudowaną autonomią i możliwością samodzielnego uczenia się, ponieważ prawa te nie mogą zostać przekształcone w kod maszynowy”, a w przypisie do tego punktu wyliczono wszystkie oryginalne prawa Asimova, łącznie z prawem zerowym.

Jak wynika z zacytowanego fragmentu, ale też jego miejsca w tekście oraz strukturze rezolucji (pkt T stanowi pierwszą spośród „Zasad ogólnych”) oraz użycia nazwy niegdyś generalnej, dziś już chyba indywidualnej

(problemy z zastosowaniami militarnymi i policyjnymi, ze zdefiniowaniem krzywdy, z ingerencją w ludzką autonomię, z relacją robot – człowiek), to przecież to, kto jest adresatem danych tu praw, nie jest bez znaczenia.

Jeśli odpowiedzialność za ich przestrzeganie ciąży na ludziach, to można rozważać, czy jest to wyłącznie odpowiedzialność moralna, czy może również prawna. Są pewne przesłanki, które przemawiają za tym, że mogłaby to być odpowiedzialność prawna. Na przykład fakt, że także zaniechanie robota i niezapobieżenie uszczerbkowi ludzkości albo człowieka miałyoby, zgodnie z prawami Asimova, należeć do zakresu zastosowania normy przewidującej tę odpowiedzialność. W etyce kwestia odpowiedzialności moralnej za zaniechanie jest mocno sporna²¹. Przeważa stanowisko, że odpowie-

21 B. Williams, *A critique of utilitarianism* (w:) J. Smart, B. Williams, *Utilitarianism for and against*, Cambridge–New York–Melbourne 1973, s. 93–150.

działność tego typu zachodzi tylko wówczas, gdy dany podmiot jest z jakiegoś tytułu obciążony obowiązkiem działania (np. strażak ma moralny obowiązek działania w celu ratowania życia ludzkiego w trakcie pożaru, podczas gdy zwykły człowiek takiego obowiązku nie ma)²². Powstaje więc pytanie, z jakiej racji (*ratio*) projektant, twórca, producent, właściciel czy inna osoba mieliby być odpowiedzialni za to, że „ich” robot nie udzielił pomocy każdemu, nawet przypadkowemu człowiekowi, żeby ten uniknął uszczerbku (krzywdy). Pomijając inne kwestie, należy zauważyć, że stworzenie i utrzymywanie robota zdolnego do takiej powszechnie niesionej pomocy może być zwyczajnie bardzo kosztowne i może stać w sprzeczności z jego zaprojektowaną funkcją (na przykład komercyjnego świadczenia usług medycznych). Co istotniejsze, ze względu na swe funkcje niektóre roboty byłyby w sposób „naturalny” bardziej zdolne do niesienia pomocy (np. roboty medyczne czy opiekuńcze). Jeśli byłoby to możliwe, robot taki winien (gdyby traktować prawo Asimova poważnie) świadczyć swą pomoc wszystkim, którzy jej potrzebują. W konsekwencji mógłby on funkcjonować jedynie jako obiekt publiczny, dostępny dla wszystkich (nieodpłatnie). Reguł tych nie sposób natomiast stosować do robotów prywatnych, z których korzystanie ma na celu osiągnięcie zysku. Roboty oparte na AI w tym zakresie nie mogą różnić się od innych narzędzi i urządzeń, tzn. do korzystania z nich będą uprawnieni tylko ci, którzy nabędą określone prawo (własność, najem itd.). A *contrario* osoby, które danego prawa do korzystania z robota mieć nie będą, nie będą mogły również skutecznie dochodzić ochrony ze strony robota do siebie nienależącego, a w każdym razie nie w zakresie większym niż jest to możliwe w odniesieniu do jakichkolwiek innych rzeczy. Powszechnie w prawie cywilnym dopuszcza się np. użycie cudzej rzeczy w stanie wyższej konieczności (np. art. 142 k.c.) celem odwrócenia niebezpieczeństwa grożącego bezpośrednio dobrom osobistym (a nawet majątkowym) danej osoby. Nie ma powodów, by w odniesieniu do robotów autonomicznych możliwość taka uległa ograniczeniu, nie wydaje się jednak możliwe stworzenie normy, która nakazywałaby nieograniczony aktywizm po stronie robota w tym zakresie.

22 J. Lucas, *Responsability*, Oxford–New York–Toronto 1993, s. 53.

Na gruncie moralnym łatwiej już można by znaleźć uzasadnienie dla obciążenia twórców i właścicieli czy innych dysponentów robota obowiązkiem zapewnienia takiego jego działania, żeby każdorazowo unikać krzywdy ludzkości (a więc zgodnie z prawem zerowym). W końcu twórcy, producenci, właściciele robotów itd. są częścią ludzkości, a jej niezakłócone funkcjonowanie leży w ich interesie. Natomiast w rygorze norm prawnych można uzasadnić odpowiedzialność za uszczerbek dla ludzkości (spowodowany czy to działaniem, czy zaniechaniem robota), na przykład poprzez odwołanie się do zasad konstytucyjnych, takich jak na przykład zasada dobra wspólnego, która tłumaczy nakładanie podatków. Albo można konstruować uzasadnienia podobne do tych, które ufundowały odpowiedzialność na zasadzie ryzyka za szkody spowodowane przez zakład czy pojazd mechaniczny poruszany siłami przyrody (tj. czerpanie korzyści ze źródła, które zwiększa poziom ryzyka w środowisku). Ale i tu powstają problemy łączące się z zakresem takiej odpowiedzialności. Jak ustalić granicę pomiędzy tymi robotami, które mają zdolność działania w danym zakresie, i tymi które nie mają. Wszak trudno od odkurzacza Roomba wymagać, żeby ratował człowieka topiącego się w wannie. Być może rozwiązanie leżałoby w odpowiednim rozłożeniu ciężaru dowodu. Wraz ze wzrastającą wszechstronnością robotów można by przejść od ciężaru dowodu, że robot powinien być działać, spoczywającego na osobie formułującej roszczenia, do ciężaru dowodu, że robot nie musiał działać, spoczywającego na podmiocie odpowiedzialnym za robota. Ale wtedy pozostają jeszcze inne problemy. Czy robot ma obowiązek reagować na ryzyko skrzywdzenia każdej osoby bez względu na jej położenie i prawa, jakie ma ona w stosunku do robota itd.? Czy też katalog takich osób należałoby przy użyciu jakiegoś kryterium ograniczyć, a jeśli tak, to według jakiego (bliskość w przestrzeni czy pod względem stosunków o charakterze prawnym, czy może wg kryterium wieku itd.)? Reagowanie przez robota na każde ryzyko uszczerbku ludzkości i na ryzyko krzywdy każdego człowieka, w konsekwencji stworzenia i utrzymywania go takim, żeby tę zdolność posiadał, jest raczej niemożliwe. Ewentualność taką należy więc na zasadzie *impossibilium nulla obligatio est* wykluczyć. A w takim razie problem ustalenia kryterium jest nieunikniony.

Pomijając jednak podniesione wątpliwości i zakładając, że miałyby zaistnieć odpowiedzialność prawna, nie uniknie się pytania o rodzaj tej odpowiedzialności. W grę wchodzi odpowiedzialność karna, cywilna, kontraktowa, deliktowa itp., ale może także jakieś nowe powstałe w przyszłości rodzaje odpowiedzialności?

Ponadto należy wziąć pod uwagę, że odpowiedzialność ludzi będzie w jakimś stopniu zależna od tego, jaki status prawny wedle nowych projektowanych regulacji z zakresu „prawa robotów” uzyskają roboty. Inaczej będzie, jeżeli pozostaną rzeczami, a inaczej, jeżeli uzna się wszystkie lub pewne z nich za obdarzone w jakimś zakresie osobowością prawną. W tym drugim przypadku można sobie wyobrazić co najmniej solidarną lub subsydiarną odpowiedzialność robota.

Nadanie choćby szczątkowej podmiotowości prawnej robotom (do rozważenia tej kwestii wzywa rezolucja PE, pkt 59 f), jeśli kiedykolwiek nastąpi, będzie konsekwencją dalszego rozwoju AI i stworzenia silnej AI, ewentualnie jakiejś postaci pośredniej między słabą a silną AI. Możliwość jej autonomicznej decyzji na wielu polach nieuchronnie prowadzić musi do postawienia pytania o zachowanie niezgodne z założeniami, którymi kierowali się twórcy. Najbardziej widoczne będzie to tam, gdzie robot będzie musiał dokonywać ocen i wyborów, których nie można z góry opisać i zaprogramować. W takim wypadku racjonalne będzie uczynienie adresatem – a ściślej: drugim adresatem – praw Asimova samego robota (zgodnie z pierwotnym konceptem). Twórca winien zaprojektować AI według określonych założeń etyczno-prawnych, jednakże w określonych granicach to robot podejmować będzie decyzję, która również musi opierać się na tych zasadach. Ich złamanie musiałoby pociągać za sobą konsekwencje w postaci jakiejś formy odpowiedzialności. Może tu chodzić nie tylko o odpowiedzialność cywilną, ale również specyficzną odpowiedzialność osobistą polegającą na wyeliminowaniu określonej AI z obrotu.

W wypadku zaawansowanej silnej AI, która uzyskałaby rodzaj podmiotowości, należałoby konsekwentnie w zakresie prawa zerowego i pierwszego oczekiwać od robota zachowania co najmniej takiego, jakiego można by na gruncie prawnym oczekiwać od człowieka. Jeśli np. normy prawne nakazują aktywną pomoc osobie znajdującej się w bezpośrednim zagrożeniu życia, to tym bardziej winny obowiązek takiej pomocy nakła-

dać na robota. Tam gdzie tego obowiązku powszechnego nie ma, nałożenie go na roboty wydaje się jednak postulatem dość idealistycznym. Dla przykładu, jeśli robot mógłby pomóc osobie znajdującej się na innym kontynencie, czy winien to zrobić zgodnie z dosłownie potraktowanym prawem robota? Czy superinteligentny robot (system AI) powinien rzeczywiście działać z korzyścią ludzkości ustaloną według sobie znanych kryteriów? Sam Asimov – twórca omawianych tu praw – zwracał uwagę na konsekwencje, do których może to doprowadzić: dla „ludzkości” największym zagrożeniem może być nie kto inny, tylko człowiek.

Już na pierwszy rzut oka widać, że prawa Asimova nie są jednolite, jeżeli chodzi o chronione dobra i podmioty, których interesów chronią.

Prawo zerowe służy ochronie ludzkości. Dotyczy sytuacji i stosunków pomiędzy osobą odpowiedzialną za robota a każdą z istot ludzkich i każdą ze wspólnot takich jednostek – łącznie z największą z nich, jaką jest ludzkość.

Prawo pierwsze służy ochronie życia i szeroko rozumianego zdrowia człowieka (w tym jego dobrostanu). Niektórzy mogą obstawiać przy tym, aby ze względów polityki prawa przyjąć szerokie rozumienie krzywdy i uznać, że prawo pierwsze chroni także prawa majątkowe człowieka. Naszym zdaniem byłaby to nadinterpretacja praw Asimova, które służą przede wszystkim bezpieczeństwu fizycznemu ludzi. Ponadto pojawiłby się problem, dlaczego chroniony miałyby być tylko majątek istot ludzkich, a już nie osób prawnych, których z pewnością prawa Asimova nie dotyczą. Prawo pierwsze dotyczy sytuacji i stosunków pomiędzy osobą odpowiedzialną za robota a każdą potencjalnie skrzywdzoną istotą ludzką.

Prawo drugie służy ochronie oczekiwań, jakie człowiek formułuje wobec robota; funkcji, jaką robot ma spełniać. Ale chyba nie można uznać, że oczekiwania wszystkich ludzi miałyby być chronione. Choćby z tej przyczyny, że wtedy mogłyby powstawać kolizje między tymi oczekiwaniami. Wydaje się, że prawo to powinno chronić wyłącznie interes osób mających z takiego czy innego tytułu prawo do dysponowania robotem lub wydawania mu poleceń.

Prawo trzecie, przynajmniej na dzisiaj, służy ochronie majątku, którego częścią jest robot, a należącego do podmiotu prawa (osoby fizycznej, prawnej). Gdyby

roboty usamodzielniały się na tyle, że w jakimś zakresie uznano by je za nie-rzeczy, na przykład przyznano by im status podobny do zwierząt albo nawet uczyniono z nich podmiot prawa, to wówczas prawo to chroniłoby też egzystencję i dobrostan samego robota, które mogłyby zostać zagrożone pod wpływem nieuprawnionych do tego podmiotów, na przykład przez człowieka, który „zhakowałby” go i wydałby mu polecenie autodestrukcji.

W związku z różnicami zachodzącym pomiędzy poszczególnymi prawami Asimova odpowiedzialność prawna za ich przestrzeganie, nawet jeżeli przypisze się ją ludziom, musi być zróżnicowana.

Dla prawa zerowego z pewnością nie będą odpowiednie metody regulacji charakterystyczne dla prawa prywatnego. Prawo cywilne nie chroni „ludzkości” jako ogółu ludzi na Ziemi i „ludzkość” nie może domagać się naprawienia szkody, którą poniosła, przy wykorzystaniu mechanizmów cywilnoprawnych. Wątpliwe jest, by istniała możliwość skonstruowania normy prawa cywilnego, która w odniesieniu do działania robota mogłaby w tym zakresie coś zmienić. W konsekwencji na płaszczyźnie prawa cywilnego prawo zerowe jest po prostu nieprzydatne. Ze względu na duże ryzyko związane z naruszeniem tego prawa właściwsze będą metody prawnokarna i prawnoadministracyjna. Odpowiedzialność taka powinna być nieograniczona w czasie, a zatem przechodzić powinna na kolejne podmioty sprawujące nadzór nad robotem albo być współdzielona na różnych zasadach. Początkowo odpowiedzialnym może być producent albo projektant, potem szeroko rozumiany serwisant software’u czy hardware’u, operator lub właściciel. Istotne znaczenie może mieć rejestracja robotów i powiązanie z nią obowiązkowego ubezpieczenia. Ze względu na ryzyko (wraz z upowszechnianiem się robotów będzie ono stale rosło) pojawienia się robotów, których nie będzie można z nikim powiązać (roboty niczyje), konieczne stanie się stworzenie funduszu gwarancyjnego obejmującego szkody wyrządzone przez tego rodzaju roboty.

W zastosowaniu do prawa pierwszego możliwe są dwa główne reżimy: prawnokarny i cywilnoprawny. W obu reżimach powodem stosowania sankcji jest naruszenie jakichś dóbr, które w języku potocznym może być określone jako „krzywda”. Jeśli więc uznać, że podstawowym celem odpowiedzialności cywilnej jest

naprawienie szkody, to krzywdzenie (o którym mowa w zerowym i pierwszym prawie Asimova) w kontekście prawa cywilnego musiałoby oznaczać nic innego jak „wyrządzenie szkody”. Krzywda (człowieka, ludzkości), która na gruncie prawa cywilnego byłaby irrelevantna, musi, jak się wydaje, pozostać taką również w prawie robotów. Trudno w praktyce wyobrazić sobie system normatywny, który dozwolone dla ludzi i niewyrządzające szkody (w ujęciu tradycyjnym) zachowania traktowałby jako bezprawne i „krzywdzące”, jeśli będą to działania maszyny. Próba wprowadzenia takiego rozróżnienia z góry skazana jest na niepowodzenie, bo szkoda (zarówno majątkowa, jak i niemajątkowa) musi być odniesiona do poszkodowanego, a nie do sprawcy czy jego zachowania. Jeśli człowiek nie poniósł żadnej szkody (rozumianej według ustalonych reguł prawa cywilnego), to nie sposób przyjąć, że został skrzywdzony – dlatego tylko, że działanie wobec niego podjął robot. Z tego punktu widzenia okazuje się, że zdekodowana na gruncie prawa cywilnego zasada „nie krzywdzić” wiąże się z ogólnymi regułami odpowiedzialności deliktowej, której rzeczywistą treść możemy zrozumieć, odwołując się do konkretnego porządku prawnego. Odwołanie się do reguł ogólnych oznaczać musiałoby zróżnicowanie zakresu ochrony w zależności od porządku wewnętrznego i oczywiście każdorazowo inaczej dekodowałoby prawa Asimova. Jeśli np. w danym porządku prawnym nie dopuszcza się naprawienia tzw. szkód moralnych albo szkód pośrednich, albo limituje się wysokość odszkodowania, to zawartość normatywna prawa Asimova byłaby inna niż tam, gdzie naprawienie takich szkód się dopuszcza. Oczywiście możliwe jest (a nawet pożądane i na gruncie europejskim prawdopodobne) skonstruowanie reguł, które będą odrębnie określać zarówno to, czym jest szkoda, jakie są przesłanki jej naprawienia, jak i to, w jakim zakresie podlega naprawieniu, jeśli jej sprawcą jest robot. Trudno jednak uznać, że będzie to w jakikolwiek sposób wynikać z praw Asimova.

Oczywiście teoretycznie „krzywda” (czy szkoda na gruncie konkretnego porządku prawnego) może oznaczać każde pogorszenie majątku czy dobrostanu człowieka (pierwsze prawo Asimova), a nawet ludzkości (prawo zerowe). Wyobraźmy sobie najprostszy przykład, z którym już dziś mamy powszechnie do czynienia. AI (w postaci niematerializowanej) dokonuje transakcji

giełdowych albo gra w gry (np. pokera) na pieniądze. Jeśli wygrywa, to przynajmniej w niektórych wypadkach doprowadza do pogorszenia stanu majątku (a niekiedy nawet i zdrowia) przeciwników (albo kontrahentów w wypadku gry giełdowej). Inny przykład: robot (tym razem AI zmaterializowana) sprząta halę magazynową. Robi to tak dobrze, że pracownik, który robił to wcześniej, traci pracę, a więc jego majątek ulegnie uszczupleniu. W obu jednak sytuacjach tak rozumiana szkoda jest dopuszczalna, bo ocena przesuwana jest na płaszczyznę oceny zachowania według reguł danego porządku prawnego (bezprawności). Zatem okoliczność, że zachowanie powoduje czyjąś uszczerbek (powoduje jakąś „krzywdę”), nie jest wystarczająca, by uznać je za niedozwolone. Konieczna jest ocena tej szkody

komunikacji poruszany za pomocą sił przyrody. Tak jak w odniesieniu do prawa zerowego, tak również i tu odpowiedzialność powinna być nieograniczona w czasie i przechodzić na kolejne podmioty sprawujące nadzór nad robotem albo być współdzielona na różnych zasadach (solidarna odpowiedzialność producenta i użytkownika, gwarancyjna odpowiedzialność zakładu ubezpieczeń, pomocnicza odpowiedzialność funduszu gwarancyjnego, serwisanta, operatora lub właściciela). W świetle tego, co już zostało powiedziane, należy uznać, że pierwsze prawo Asimova musi być ostatecznie sprowadzone do reguł wykształconych na gruncie cywilistyki. Zakaz krzywdzenia człowieka oznaczać będzie nic innego jak znaną od dawna zasadę *nemi-nem laedere* i zasadę odpowiedzialności za wyrządzoną



Nie ma powodów, by w zakresie odpowiedzialności za szkody wyrządzone przez robota przyjmować *a priori* absolutną odpowiedzialność niezależną od takich okoliczności jak wina czy przyczynienie się poszkodowanego.

przez pryzmat reguł określonego porządku prawnego, które – co charakterystyczne – w wielu wypadkach granicznych odwołują się do ocen nieostrych, pozaprawnych systemów aksjologicznych, reguł etyki itd. W konsekwencji można postawić pytanie, czy zerowe i pierwsze prawa Asimova zawierają jakąś inną treść i powinny być jakoś inaczej interpretowane niż nakaz nieszkodzenia kierowany do innych podmiotów.

Jeżeli mimo powyższych zastrzeżeń chcemy pierwsze prawo Asimova wtłoczyć w reżim cywilnoprawny, to w grę wchodzi na przykład odpowiedzialność analogiczna do dotychczas funkcjonującej odpowiedzialności na zasadach ogólnych za szkodę majątkową i niemajątkową, odpowiedzialności za szkodę wyrządzoną przez produkt niebezpieczny, odpowiedzialności za szkodę wyrządzoną przez zwierzęta, odpowiedzialności za szkodę spowodowaną przez mechaniczny środek

szkodę. Szkoda ta może obejmować zarówno dobra osobiste człowieka, jak i jego majątek – w tym oczywiście również inne roboty. Przesłanki tej odpowiedzialności zapewne (tak jak w wypadku szkód wyrządzonych przez produkt niebezpieczny, ruch przedsiębiorstwa czy pojazdów mechanicznych) muszą abstrahować od elementów subiektywnych (tj. winy), choć ze względu na autonomiczność robota (zwłaszcza sterowanego przez silną AI) nie jest to kwestia oczywista. Nie ma powodów, by w zakresie odpowiedzialności za szkody wyrządzone przez robota przyjmować *a priori* absolutną odpowiedzialność niezależną od takich okoliczności jak wina czy przyczynienie się poszkodowanego. Zwłaszcza że na gruncie prawa w ogóle pojęcie winy (podobnie jak pojęcie związku przyczynowego) i tak będzie prawdopodobnie wymagało przemyślenia, w miarę jak w życiu społecznym i obrocie prawnym coraz częściej i w coraz

większym zakresie będą uczestniczyć roboty. Pierwsze prawo Asimova nie może być też wskazówką dla prawa cywilnego w zakresie określenia podmiotu, przeciwko któremu nie może zwrócić się robot. Jest zupełnie oczywiste, że naprawienia szkody mógłby dochodzić również inny niż człowiek podmiot prawa cywilnego (osoba prawna). W wypadku nadania robotom jakiejś postaci podmiotowości – pokrzywdzonym może być inny robot, a co za tym idzie nie można wykluczyć sporów sądowych między robotami.

Respektowaniu prawa drugiego sprzyjałaby w największym stopniu odpowiedzialność cywilna za nie należyte wykonanie zobowiązania lub z tytułu rękojmi czy gwarancji. W przypadku zachowania robota polegającego wyłącznie na niewykonaniu polecenia człowieka (przy niewystąpieniu szkody) trudno znaleźć uzasadnienie dla jakichś innych reżimów odpowiedzialności. Przedmiotem dyskusji może stać się to, czy taka odpowiedzialność powinna być ograniczona w czasie, ewentualnie jak długo powinna trwać. Bardzo trudne będzie oczywiście zrealizowanie (od strony technicznej) postulatu, by prawo drugie nie stało w sprzeczności z pierwszym. Nakaz wykonywania poleceń człowieka wchodzący w kolizję z nakazem niewyrządzania człowiekowi krzywdy mogą z punktu widzenia odpowiedzialności cywilnej rodzić jeszcze większe trudności. Czy robot może wartościować krzywdę (dobrostan) i oceniać, które polecenia wykonać, a które nie, albo porównywać szkodę swojego dysponenta, która wyniknie z niewykonania polecenia, ze szkodą, która powstanie, gdy polecenie zostanie wykonane? *Prima facie* problemy, które się z tym wiążą, wydają się po prostu nierozwiązywalne. Jeśli przyjmiemy, że polecenia realizują jakiś interes uprawnionego dysponenta robota, to niewykonanie polecenia może wiązać się (a nawet zwykle będzie się wiązać) z naruszeniem tego interesu, a zatem jakąś postacią szkody dla dysponenta. Z pewnością możliwe jest stworzenie systemów zabezpieczających przed wykonaniem polecenia, które skutkowałoby wyrządzaniem na zewnątrz szkód dających się określić jako znaczące czy oczywiste (np. zabicie czy ciężkie uszkodzenie ciała człowieka). Jednakże tam, gdzie chodzi o szkodę mniej oczywistą dla zewnętrznego obserwatora (np. związaną z przekroczeniem granic swobody wypowiedzi albo szkodą w postaci *lucrum cessans*), wydaje się

niemożliwe stworzenie systemu, który dokonywałby oceny, czy należy wykonać polecenie człowieka, czy też raczej nie wykonać, a w konsekwencji spowodować u niego jakiś uszczerbek. Ilość możliwych następstw (zwłaszcza pośrednich) dowolnego zachowania jest nieograniczona, wobec czego analiza potencjalnych szkód wyrządzonych przez robota przy wykonaniu polecenia musiałaby być w jakiś sposób limitowana.

W stosunku do prawa trzeciego właściwa byłaby wyłącznie odpowiedzialność cywilna analogiczna do odpowiedzialności z tytułu rękojmi lub gwarancji. Przedmiotem dyskusji może stać się to, czy taka odpowiedzialność powinna być ograniczona w czasie, ewentualnie jak długo powinna trwać. Problemy opisane w odniesieniu do drugiego prawa stają się tu jeszcze bardziej wyraźne. Robot miałby dokonywać oceny wielokierunkowej, odnoszącej się do zakresu potencjalnych szkód, które mogłyby przez różne jego zachowania (lub zaniechania) zostać spowodowane. Zastosowanie trzeciego prawa robotyki musiałoby prowadzić do wniosku, że robot nie może podejmować żadnych albo prawie żadnych skutecznych działań obronnych w stosunku do ludzi, większość z nich bowiem może wiązać się z jakąś krzywdą człowieka. W praktyce można podejść do zagadnienia trzeciego prawa Asimova na dwa sposoby. Pierwszym byłoby uznanie, że prawo to nie obowiązuje w relacjach robot – człowiek, tzn. robot w ogóle nie może bronić się przed zagrożeniem ze strony człowieka. Z punktu widzenia lęków przed wszechmocnymi maszynami taka koncepcja wydaje się dość atrakcyjna. Drugim możliwym rozwiązaniem byłoby dopuszczenie w tych relacjach jakiegoś rodzaju obrony koniecznej robota. To pozornie racjonalne podejście grozi jednak naruszeniem godności człowieka, który mógłby znaleźć się w gorszej pozycji niż maszyna. W dłuższej perspektywie, gdyby siły (fizyczne i intelektualne) robotów znacząco przewyższyły potencjał ludzki, oznaczałoby to groźbę, że świat stanie się światem maszyn, którym – mówiąc obrazowo – ludzie powinni grzecznie się kłaniać, by uniknąć „dyscyplinującego” działania robota. Bardziej racjonalne – z ludzkiej perspektywy – jest zatem konsekwentne traktowanie robota jako rzeczy, która nie może – w żadnych okolicznościach – narzucać swej woli człowiekowi (zatem i aktywnie się bronić). Jeśli jednak dopuści się jakąś postać personifikacji AI, zwłaszcza

w wypadku silnej AI, a tym bardziej samoświadomej, to i dopuszczenie samoobrony takiego podmiotu wydaje się nieuniknione. W takim wypadku jednak trudno by z kolei było wskazać, dlaczego robota miałyby ograniczać trzecie prawo Asimova, tzn. dlaczego miałyby on poświęcać swą podmiotowość w imię kaprysu człowieka.

Oprócz powyższego należy pamiętać, że za każdym razem gdy naruszenie praw nastąpiłoby z powodu ingerencji jakiegoś podmiotu nieuprawnionego (który na przykład „zhakowałby” robota, bezpośredniego sprawcę naruszenia któregoś z praw) w grę wchodziłyby odpowiedzialność cywilna i karna tego podmiotu, obok ewentualnej odpowiedzialności karnej lub cywilnej podmiotu sprawującego nadzór nad robotem za niedopełnienie obowiązków.

Gdyby zaś szeroko rozumianym robotom przyznano w jakimś zakresie osobowość prawną, w grę wchodziłyby także samodzielna, solidarna lub subsydiarna odpowiedzialność cywilna takiego nowego podmiotu prawa, ale także odpowiedzialność karna, na przykład analogiczna do odpowiedzialności podmiotów zbioro-

4. Podsumowanie

Powyższa analiza wykazuje, że prawa Asimova, nawet jeżeli za ich adresatów uznaje się ludzi sprawujących kontrolę nad robotem, nie są wystarczającą teoretyczną podstawą dla budowania konkretnej regulacji prawnej, zwłaszcza z zakresu prawa cywilnego. Co najmniej (abstrahując od innych braków opisanych powyżej), nie obejmują swym zakresem wszystkich sytuacji wymagających uregulowania. Na przykład w ogóle nie odnoszą się do kwestii odpowiedzialności za działania lub zaniechania robota wobec podmiotów prawa niebędących ludźmi. Wątpliwe jest też ustanowienie prostej i bezwarunkowej odpowiedzialności za brak działania robota, w sytuacji gdy mogłoby ono zapobiec uszczerbkowi ludzkości lub krzywdzie człowieka. Ponadto prawa Asimova nie różnicują robotów ze względu na istotne ich cechy, które powinny wpływać na wiążącą się z nimi odpowiedzialność. Wszystko to wskazuje, że przywołanie praw Asimova w rzeczowej rezolucji PE nie ma w żadnym razie charakteru normatywnego. Można co najwyżej uznać, że stanowi



Racjonalne – z ludzkiej perspektywy – jest konsekwentne traktowanie robota jako rzeczy, która nie może – w żadnych okolicznościach – narzucać swej woli człowiekowi lub aktywnie się bronić.

wych za czyny zabronione pod groźbą kary. Przyznanie osobowości podważałoby jednak – co było już sygnalizowane – sensowność drugiego prawa Asimova, tzn. obowiązku wykonywania poleceń człowieka. Daleko posunięta autonomiczność, wynikająca z wysokiej, wieloaspektowej inteligencji, połączona z dużymi możliwościami wpływania na rzeczywistość fizyczną (robot wielozadaniowy, android) musiałaby wiązać się z koniecznością subiektywizacji odpowiedzialności, tj. analizowania określonych zachowań nie tylko przez pryzmat zachowania określonych reguł postępowania, ale – jak w wypadku człowieka – również odniesienia tego do przesłanek subiektywnych, takich jak wina.

ono bardzo ogólne i niekoniecznie wiążące wskazanie pewnych wartości kulturowych i moralnych, których ochrona i realizacja są przez Parlament Europejski zalecane. Ale ponieważ podobne lub te same wartości są wskazywane również w innych miejscach tej rezolucji, to przywołanie praw Asimova można uznać za redundantne – mimo pozorów obiektywizmu odwołujące się tak naprawdę do emocji lub będące jedynie retorycznym środkiem estetycznym bądź perswazyjnym.

Ten stan rzeczy daje dobre powody, by przypuszczać, że przyczyną przywołania praw Asimova w rezolucji nie były jakieś głębokie refleksje autorów tego dokumentu, ale raczej trwający wciąż jeszcze w kulturze

stan zawieszenia pomiędzy dwoma paradygmatami robotyki: tradycyjnym paradygmatem androidalnym (grec. ἀνδρικός – męczyzna, istota ludzka), wedle którego robot ma kształt człowieka i jest jego analogonem, a współczesnym paradygmatem neuronalnym, wedle którego postać fizyczna robota nie ma w gruncie rzeczy znaczenia, najważniejsze bowiem jest jego uposażenie kognitywne. Dopóki więc nowy paradygmat nie zwycięży w umysłach ludzi i podręczniki nie zostaną ostatecznie napisane na nowo²³ ryzyko takich nieporozumień wciąż będzie trwało.

Artykuł powstał dzięki wsparciu Narodowego Centrum Nauki, grant 2018/29/B/HS5/00421.

Bibliografia

- Anderson S.L., *Asimov's „Three Laws of Robotics” and Machine Metaethics*, „AI & Society” 2008, vol. 22, nr 4, s. 477–493.
- Asimov I., *Bicentennial Man*, Ballantine Books, New York 1976.
- Asimov I., *Robots and Empire*, Doubleday Books, New York 1985.
- Asimov I., *Runaround*, Astounding Science Fiction, New York 1942.
- Barbrook R., *Imaginary Futures. From Thinking Machines to the Global Village*, Pluto Press, London 2007.
- Bostrom N., *Superintelligence. Paths, Dangers, Strategies*, Oxford University Press, Oxford 2014.
- Clarke R., *Asimov's laws of robotics: Implications for information technology. Part 2, „Computer”* 1994, vol. 27, nr 1, s. 57–66.
- Commission Staff Working Document, *Liability for emerging digital technologies*, COM(2018) 237 final, Brussels, 24.4.2018, <https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX%3A52018SC0137>.
- EC High-Level Expert Group on AI, *A definition of Artificial Intelligence: main capabilities and scientific disciplines*, Brussels (8.4.2019), <https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines>. (dostęp 17.06.2020).
- EC High-Level Expert Group on AI. (8.4.2019), *Ethics Guidelines for Trustworthy AI*, <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>. (dostęp 17.06.2020).
- EC Joint Research Centre, *Artificial Intelligence. A European Perspective*, Luxembourg 2018. <https://publications.jrc.ec.europa.eu/repository/bitstream/JRC113826/ai-flagship-report-online.pdf> (dostęp 22.06.2020).
- European Commission, *Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics*, Brussels, 19.2.2020 COM (2020) 64 final, https://ec.europa.eu/info/sites/info/files/report-safety-liability-artificial-intelligence-feb2020_en_1.pdf.
- Expert Group on Liability and New Technologies – New Technologies Formation, *Liability for Artificial Intelligence and Other Emerging Digital Technologies*, European Union 2019.
- Księżak P., *Zdolność prawna sztucznej inteligencji (w:) Czynyć postęp w prawie. Księga jubileuszowa dedykowana Profesor Birucie Lewaszkiewicz-Petrykowskiej*, red. W. Robaczyński, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2017, s. 63–72.
- Kuhn T., *The Structure of Scientific Revolutions*, The University of Chicago Press, Chicago–London 1996.
- Lucas J., *Responsability*, Oxford University Press, Oxford–New York–Toronto 1993.
- McCauley L., *AI Armageddon and the Three Laws of Robotics*, „Ethics and Information Technology” 2007, vol. 9, nr 2, s. 153–164.
- Murphy R., Woods D., *Beyond Asimov: The Three Laws of Responsible Robotics*, „IEEE Intelligent Systems” 2009, vol 24, nr 4, s. 14–20.
- Nilsson N., *The Quest for Artificial Intelligence: A History of Ideas and Achievements*, Cambridge University Press, New York 2009.
- Rezolucja Parlamentu Europejskiego z dnia 16 lutego 2017 r. zawierająca zalecenia dla Komisji w sprawie przepisów prawa cywilnego dotyczących robotyki, P8_TA(2017)0051 (Dz.Urz. UE C 252 z 18.7.2018, s. 239–257).
- Searle J.R., *Minds, Brains and Science*, Harvard University Press, Cambridge Ma 1984.
- Weng Y.-H., Chen C.-H., Sun C.-T., *Safety Intelligence and Legal Machine Language: Do We Need the Three Laws of Robotics?* (w:) *Service Robot Applications*, ed. Y. Takahashi, Intech-Open 2008, s. 195–214, https://www.intechopen.com/books/service_robot_applications/safety_intelligence_and_legal_machine_language_do_we_need_the_three_laws_of_robotics (dostęp 17.06.2020).
- Williams B., *A critique of utilitarianism* (w:) Smart J., Williams B., *Utilitarianism for and against*, Cambridge University Press, Cambridge–New York–Melbourne 1973, s. 77–150.

23 T. Kuhn pisał, że zwycięstwo nowego paradygmatu nie będzie pełne, jeżeli nie zostaną napisane od nowa podręczniki. Podręczniki pełnią bowiem niezwykle istotną rolę w internalizowaniu paradygmatu nie tylko przez szeroką publiczność czy nowicjuszy w danej dziedzinie, ale wszystkich przedstawicieli tzw. nauki normalnej. T. Kuhn, *The Structure of Scientific Revolutions*, The University of Chicago Press, Chicago–London 1996, s. 149.