Zoon Politikon 13 / 2022
http://www.ejournals.eu/Zoon-Politikon/
DOI: 10.4467/2543408XZOP.22.008.16585

Prawa potencjalnych cyfrowych osób w cyfrowym społeczeństwie.
Kilka uwag o moralno-prawnym statusie ultrainteligentnej maszyny

Andrzej Stoiński*

Abstrakt

Zasadniczym celem tego tekstu jest zbadanie możliwości posiadania praw przez inteligentną maszynę. Zastanowimy się nad wartością trzech rodzajów uzasadnień owych praw. Po pierwsze, uzasadnienie praw sztucznej inteligencji odwołujące się do ich korelacji z ludzkimi obowiązkami niezupełnymi. Analizujemy też podejście bazujące na statusie obiektów moralnych w ich relacjach z ludźmi. Powołując się na prace różnych teoretyków oba te uzasadnienia uznajemy za niewystarczająco przekonujące. Najbardziej trafne wydaje się uzasadnienie posiadania praw przez takie maszyny, które wypełniałyby warunki bycia sztuczną osobą, moralnym podmiotem. Jednakże i ta ostatnia opcja nie jest wolna od licznych trudności, na które zwracamy uwagę.

Słowa kluczowe: *sztuczna inteligencja, SI, etyka sztucznej inteligencji, sztuczna osoba, teorie praw, prawa/obowiązki sztucznej inteligencji*

Rights of Possible Digital Persons in a Digital Society.
A few remarks on the moral and legal status of an ultra-intelligent machine

Abstract

The main goal of this text is to take a closer look at the possibility of an intelligent machine having rights. First of all, we will consider the value of the three kinds of justifications for these rights. As for the justification of the rights of Artificial Intelligence referring to some correlation with human imperfect duties and that which is based on the status of moral objects in relations with people, we consider them unconvincing. It seems most accurate to justify the possession of rights by such machines that would fulfill the conditions of being an artificial person, a moral subject. However, the latter option is not free from the numerous difficulties that we bring to your attention.

Keywords: *artificial intelligence, AI, ethics of artificial intelligence, artificial person, theories of rights, rights/obligations of artificial intelligence*

* Dr Andrzej Stoiński, Instytut Filozofii, Uniwersytet Warmińsko-Mazurski w Olsztynie, Polska
e-mail: andrzej.stoinski@uwm.edu.pl | ORCID: https://orcid.org/0000-0002-3104-1265

Data wpłynięcia tekstu do redakcji: 28 lutego 2022 r.

Wstęp

Wraz z rozwojem technologii cyfrowych doskonalić się będą także formy sztucznej inteligencji (SI). Całkiem prawdopodobny jest moment gdy którąś z nich rozpoznamy jako superinteligentną istotę. Jednakże zetknięcie się z nią w takiej postaci wcale nie musi być jednoznaczne ze spotkaniem kogoś, komu winny przysługiwać prawa człowieka. Zasadniczym celem tego tekstu jest bliższe przyjrzenie się możliwości posiadania praw przez inteligentną maszynę. Przede wszystkim zastanowimy się nad wartością trzech rodzajów uzasadnień owych praw. Główne tezy dotyczą wartości owych uzasadnień. Twierdzimy, że uzasadnienie powołujące się na ludzkie obowiązki niezupełne wobec inteligentnych maszyn nie wystarcza do uznania, że winny im przysługiwać skorelowane z tymi obowiązkami prawa. Uważamy również, że uzasadnienie praw statusem obiektów moralnych jako uczestników relacji z ludźmi także nie jest wystarczająco przekonujące. Najbardziej trafne wydaje się uzasadnienie posiadania praw przez takie maszyny, których charakterystyka byłaby analogiczna do ludzkich osób, a tym samym wypełniałyby one warunki bycia sztuczną osobą. Jednakże i ta koncepcja nie jest wolna o trudności.

Z powyższymi celami i tezami wiążą się pytania badawcze. Dotyczą one między innymi tego: czym jest sztuczna inteligencja? Czym są prawa podmiotowe i komu one przysługują? Jak próbuje się uzasadnić prawa SI? Czym winien charakteryzować się podmiot posiadający prawa? Jakie przeszkody stoją na drodze do przyznanie maszynom praw?

Próbom odpowiedzi na te kwestie podporządkowana jest struktura tekstu. Jest on podzielony na kilka sekcji. Zaczyna się od szkicowej charakterystyki sztucznej inteligencji oraz elementarnych informacji o tym, czym są prawa podmiotowe. Główną część pracy poświęcimy uzasadnieniu praw SI. Przyjrzymy się propozycjom potraktowania jej jako

obiektu moralnego, a następnie jako moralnego podmiotu. Przeanalizujemy również warunki konieczne dla uznania SI za sztuczną osobę wraz z przysługującymi osobom uprawnieniami. Na koniec zamieścimy wybrane głosy krytyczne wobec praw SI.

Sztuczna inteligencja

Rozważania zaczniemy od doprecyzowania przedmiotu. Pomimo braku powszechnie akceptowanej jednej definicji „inteligencji”[1], co odnosi się także do „sztucznej inteligencji”[2], to przyjmujemy za Tomaszem Zalewskim, że SI to „[...] system, który pozwala na wykonywanie zadań wymagających procesu uczenia się i uwzględniania nowych okoliczności w toku rozwiązywania danego problemu i który może w różnym stopniu – w zależności od konfiguracji – działać autonomicznie oraz wchodzić w interakcję z otoczeniem” (Zalewski 2020: 3).

Takie ogólne określenie oprócz innych zalet ma także tę, że nie koliduje z klasyfikacją rodzajów SI według ich technologicznego zaawansowania[3]. W tym względzie wyróżnia się: sztuczną inteligencję w sensie

[1] Shane Legg i Marcus Hutter przytaczają 70 różnych definicji inteligencji podzielonych na trzy kategorie: encyklopedyczne, psychologiczne, informatyczne. (Legg, Hutter 2007: 17-24).

[2] Jak pisze Paul Dumouchel: „Sztuczna inteligencja, podobnie jak ludzka, nie jest kategorią dobrze zdefiniowaną. Nie odpowiada jednej zdolności, ale zbiorowi technologii obliczeniowych inspirowanych niektórymi ludzkimi zdolnościami poznawczymi” (Dumouchel 2019: 243). Podobnie na ten temat uważa (Copeland 2000). Co więcej, „[...] nie ma czegoś takiego jak uniwersalna skala, według której moglibyśmy porównywać ludzką i sztuczną inteligencję” (Dumouchel 2019: 244). Poza tym, jak podkreśla Ted Peters, należałoby też rozstrzygnąć: czy, ewentualnie, którego rodzaju sztuczny twór może w ogóle być inteligentny, w takim rozumieniu tego słowa, w jakim odnosimy je do ludzi (Peters 2019: 261). W praktyce termin „sztuczna inteligencja” nadawany jest: nauce, jednej z dziedzin informatyki, systemowi, technologii, potencjalnej myślącej maszynie, samouczącym i samodoskonalącym się narzędziom, zbiorowi metod itp. Jednakże najczęściej pojęcie to odnoszone jest do „[...] systemów, które wykazują inteligentne zachowanie dzięki analizie otoczenia i podejmowaniu działań – do pewnego stopnia autonomicznie – w celu osiągnięcia konkretnych celów” (*Sztuczna inteligencja dla Europy* 2018).

[3] Poniższa klasyfikacja zaczerpnięta od (Hassani, Silva, Unger, Mazinani, Mac Feely 2020: 146). Podobny podział na: AI, AGI oraz Superintligencję patrz (Bostrom, Yudkowsky 2011).

słabym[4] (*artificial narrow intelligence* – ANI)[5]; ogólną sztuczną inteligencję (*artificial general intelligence* – AGI)[6] i sztuczną superinteligencję (*artificial superintelligence* – ASI)[7]. Na dalszych stronach pod terminem „sztuczna inteligencja" będziemy rozumieć byt, który odpowiada charakterystyce AGI lub ASI.

Prawa podmiotowe

Najogólniej rzecz biorąc, prawa są tytułami bądź roszczeniami do pewnych benefitów lub do tego, by inne podmioty nie ingerowały w sferę zastrzeżoną dla uprawnionego. Nie ma tu potrzeby, by szerzej rozpisywać

[4] „Twierdzenie, że maszyny mogą działać inteligentnie (lub, być może lepiej, zachowywać się tak, jakby były inteligentne) nazywane jest przez filozofów hipotezą «słabej sztucznej inteligencji», a twierdzenie, że maszyny, które to robią, faktycznie myślą (w przeciwieństwie do symulowania myślenia) nazywa się hipotezą «silnej sztucznej inteligencji»" (Russell, Norvig 2003: 947). Searle pisze, że „Z punktu widzenia silnej sztucznej inteligencji, odpowiednio zaprogramowany komputer nie tylko symuluje posiadanie umysłu; on dosłownie ma umysł" (Searle 2004: 66).

[5] Są to właściwie raczej sztuczne inteligencje w postaci narzędzi służących np. do rozpoznawania twarzy, jazdy samochodem, optymalizacji procesów przemysłowych, diagnostyki chorób, tłumaczenia języków naturalnych itp. Obecne sztuczne inteligencje (słabe) są częścią systemów, nie są indywidualnościami, są zdolnymi do funkcjonowania jednocześnie w wielu miejscach matematycznymi obiektami uzależnionymi w swoim działaniu od innych elementów.

[6] Jest to postulowana sztuczna inteligencja w sensie silnym, czyli maszyna posiadająca wszystkie nasze umysłowe właściwości: uczenia się, postrzegania i rozumienia. AGI to „interaktywna, autonomiczna, samoucząca się jednostka (*agency*), która umożliwia artefaktom obliczeniowym wykonywanie zadań, które w innym przypadku wymagałyby adekwatnego wykorzystania ludzkiej inteligencji" (Taddeo, Floridi 2018: 751).

[7] To ultrainteligentna maszyna ze zdolnościami kognitywnymi rozwiniętymi w wyższym stopniu niż u człowieka. To pojęcie używane jest do opisu pewnego potencjalnego bytu zdolnego do ciągłego uczenia się, naukowego odkrywania i wykorzystywania tych informacji do konstruowania nowych narzędzi służących wytwarzaniu dóbr służących człowiekowi. Byłaby to inteligentniejsza od ludzi maszyna (program) lub ich sieć, zdolna do stałego samodoskonalenia. Na temat ultrainteligentnej maszymy klasyczny już dziś tekst: *Speculations Concerning the First Ultraintelligent Machine* (Good 1965). Zdaniem Irvinga Gooda powstanie takiej pierwszej ultrainteligentnej maszyny będzie osiągnięciem, po którym człowiek nie będzie już niczego potrzebował wynajdywać (ibidem 33).

się o pochodzeniu czy charakterystyce praw podmiotowych[8]. Należy wszakże wskazać, że pierwotnie wywodzą się one z obiektywnych uprawnień jako przedmiotu sprawiedliwości, które z czasem przybrały postać podmiotowych (subiektywnych) praw naturalnych, a współcześnie tak zwanych praw człowieka.

W literaturze przedmiotu uzasadnienie praw ma na ogół charakter fundacjonistyczny lub funkcjonalistyczny. Zgodnie z fundacjonizmem przysługują one osobom jako szczególnym bytom, przynajmniej potencjalnie samoświadomym i racjonalnym oraz zdolnym do rozpoznania dobra i zła i działania w tych kategoriach. Fundacjonistyczne podejście zakłada więc, że prawa przynależą bytom o pewnych właściwościach, często też łączy ono prawa osoby z jej godnością (Tasioulas 2012: 26-43).

W ramach podejścia funkcjonalistycznego zakłada się, że prawem jest coś, co jak pisze Charles Beitz, wiąże się ze szczególną rolą, jaką owe prawa odgrywają (Beitz 2009: 102). Dwa dominujące nurty funkcjonalizmu odwołują się odpowiednio: do wyrażania *woli* bądź *interesów* podmiotów[9]. Zgodnie z teorią woli czy też wyboru (*will, choice theory*)[10], prawami są tylko roszczenia artykułujące podmiotowy wybór[11]. W ramach teorii korzyści bądź interesu (*benefit, interest theory*)[12] termin „prawa"

[8] O ewolucji praw przedmiotowych w prawa podmiotowe, a następnie w prawa człowieka zob.: (Strauss 1953); (Tuck 1979); (Villey 1983); (Tierney 1997; Tierney 2004); (Cranston 1983); (Marshall 1950: 71-72, 84); (Buchanan 2010).

[9] Na temat historii oraz współczesnego stanu debaty o uprawnieniach w perspektywie teorii *woli i korzyści* zob. (Kramer, Simmonds, Steiner 1998).

[10] Na równoznaczność tych nazw zwraca uwagę (Harel 2005: 193). Jako inspiratorów oraz rzeczników tej koncepcji można wymienić: Immanuela Kanta, Carla von Savigny'ego, Bernharda Windscheida, Herberta Harta, Hansa Kelsena, Carla Wellmana, Hillela Steinera.

[11] Leif Wenar twierdzi, że zwolennicy *Will Theory* określają funkcję praw jako przydzielanie sfer wolności. (Wenar 2005: 223).

[12] Wśród jej orędowników da się wskazać: Jeremy'ego Benthama, Rudolfa von Jheringa, Johna Austina, Davida Lyonsa, Neila MacCormicka, Josepha Raza. Według Wenara zwolennicy koncepcji interesu uważają prawa za służące obronie dobrostanu (*well-being*) (ibidem). Koncepcja ta nosi również miano „benefit theory" (Harel 2005: 193).

odnosi się do roszczeń wyrażających interesy podmiotu (Wenar 2005: 238).

Ostatnią z tych koncepcji pominiemy jako niezbyt przydatną dla podejmowanych tu zagadnień. Idzie głównie o to, że kategoria korzyści może być rozumiana tak szeroko, że da się ją zastosować dla zbyt wielu obiektów, niekoniecznie nawet świadomych, inteligentnych, czy choćby żywych. Natomiast wypełnienie przez sztuczną inteligencję warunków właściwych teorii wyboru (na przykład takich, jak autonomia decyzyjna, która mogłaby wyrażać się w działaniem wbrew wpisanym komendom, regułom) jednocześnie potwierdza istnienie podmiotu (osoby), który także zgodnie z fundacjonistycznym uzasadnieniem winien posiadać prawa.

Zgodnie ze schematem Wesley'a Newcomba Hohfelda, prawa są skorelowane z obowiązkami, a każdy z tych fenomenów występuje w jednej z czterech form (przypadków, typów). Przypadkowi (formie) prawa jednego podmiotu nieuchronnie towarzyszy określony typ obowiązku stojący po stronie kogoś innego[13]. Nawiązując do tego schematu można powiedzieć, że bardziej ogólnie wzięte prawo do życia (jako zespół Hohfeldiańskich przypadków) koreluje z obowiązkiem (jako konglomeratem konkretnych form) znajdującym się, w jakimś miejscu spektrum rozciągającego się od negatywnego obowiązku nienaruszania życia innych do pozytywnego obowiązku jego ochrony i wspierania[14].

[13] Korelacje mają następujące postaci:
- roszczenie (*right, claim-right*) – obowiązek (*duty*);
- przywilej (wolność) (*privilege*) – brak roszczenia (*no-claim*);
- władza (kompetencja) (*power*) – podległość władzy (kompetencji) (*liability*);
- immunitet (*the immunity*) – brak władzy (kompetencji) (*disability*) (Hohfeld 1917: 710).

[14] W literaturze przedmiotu wspomina się w związku z tym o dwu-, trój-, a nawet czwórdzielnnych typologiach obowiązków. Typologia: dwódzielna patrz na przykład (Waldron 1989: 511). O trójdzielnej (*tripartite typology*) – (Shue 1996: 52). Czwórdzielną (*quadripartite typology*) niepowodowania krzywdy lub zła; promowania lub urzeczywistniania dobra; ochrony przed złem lub krzywdą; eliminowania zła, proponuje (Frankena 1973: 47).

Uzasadnienie praw SI

Uczestnicy relacji o charakterze moralnym

Skorelowanie praw jednego podmiotu z obowiązkami innego oznacza istnienie jakiegoś stosunku między nimi o charakterze etycznym, a być może także jurydycznym. W interesujących nas relacjach możemy wyróżnić podmiot[15] działania moralnego[16], akt i przedmiot, na który ów akt jest nakierowany. Przedmiot, czyli obiekt moralny (*moral patient*)[17] to byt, na który może zostać nakierowane etycznie nacechowane działanie (czyli powodujące dobro lub zło[18]) ze strony podmiotu moralnego (*moral agent*). Zakładamy, że każdy podmiot moralny jest też moralnym obiektem, choć oczywiście nie każdy obiekt moralny musi być zarazem moralnym podmiotem.

Obiekty moralne, to byty wobec których możemy mieć obowiązki, nawet jeśli one nie mają ich wobec nas. Niektórzy uważają, że SI będzie obiektem moralnym jeśli będzie miała takie własności, jak zdolność do odczuwania cierpienia[19], czy wyższą inteligencję (Bostrom, Yudkowsky

[15] Pisząc o podmiocie mamy na myśli to, co w literaturze przedmiotu określa się mianem „agenta", a co jak precyzuje Marlena Jankowska jest „pojęciem opisującym substrat, poprzez który działa AI" i który może być rozumiany jako: robot lub system ekspertowy lub software (Janowska 2015: 177-178).

[16] Podział na podmiot etyczny implicite (*implicit ethical agent*) i podmiot etyczny eksplicite (*explicit ethical agent*) (zob. Moor 2006: 19-20). Te pierwsze posiadają oprogramowanie (wpisany kod), które implicite wspiera zachowania uznawane za etyczne. Drugie, przeprowadzają analizy i działają w oparciu o kategorie etyczne oraz posiadają zdolność do uzasadnienia i wyjaśnienia swoich sądów.

[17] Terminami *patient* oraz *agent* w takim kontekście posługuje się między innymi (Floridi 2013: 135-136).

[18] Jeśli spojrzymy w tym kontekście na sytuację chociażby małych dzieci, czy osób niepełnosprawnych umysłowo, to okaże się, że jak na to zwraca uwagę Kamil Mamak: „Aby stać się ofiarą przestępstwa, nie trzeba być podmiotem moralnym" (Mamak 2022: 4).

[19] Robert Sparrow uważa, że dopóki nie da się powiedzieć o maszynach, że cierpią, dopóty nie mogą one być przedmiotami moralnej troski (Sparrow 2004: 204). Zdolność do cierpienia jako uzasadnienie moralnego statusu, patrz (Coeckelbergh 2012: 13). David Gunkel rozważa zdolność do odczuwania cierpienia już jako warunek posiadania praw, a zarazem ustanowienia podmiotu moralnego (Gunkel 2018a: 92).

2011: 7). Wydaje się, że w takich okolicznościach wobec SI jakieś obowiązki moralne winniśmy mieć. Gdyby ponadto okazało się, że byty te są samoświadomymi[20] podmiotami moralnymi, osobami, a tym samym posiadaczami praw człowieka[21], to znaczyłoby, że mamy wobec nich, skorelowane z ich prawami, także obowiązki legalne.

Rodzaje uzasadnień praw SI

Co do uzasadnienia posiadania praw przez maszyny to rozpatrzymy trzy podejścia wspomniane przez Johna-Stewarta Gordona i Svena Nyholma, są to: (1) uzasadnienie fundacjonistyczne, w którym prawa przynależą po Kantowsku charakteryzowanemu autonomicznemu bytowi; (2) uzasadnienie oparte na posiadaniu wobec maszyn obowiązków niezupełnych (formułowane np. przez Darling); (3) uzasadnienie podejściem relacyjnym (np. Coeckelbergha czy Gunkela) – gdzie o moralnym statusie decyduje specyfika związku pomiędzy człowiekiem i robotem (Gordon, Nyholm 2021). W dużym uproszczeniu można napisać, że w przypadku (1) SI byłaby moralnym podmiotem, a zarazem podmiotem praw, w opcji (2) i (3) SI mogłaby owszem być podmiotem praw, ale byłaby tylko moralnym obiektem. Spośród tych trzech rodzajów uzasadnień bliżej przyjrzymy się (1), które z tej racji zaprezentujemy po krótkim omówieniu (2) i (3).

---

[20] Badaniami nad potencjalną świadomością SI zajmuje się między innymi Susan Schneider. Argumentuje ona, że jeśli SI jest świadoma, to winna jej przysługiwać ta sama legalna ochrona, jaka przynależy innym świadomym istotom (Schneider 2020: 454-455). Kluczowa dla tego rodzaju testów byłaby odpowiedź na pytanie, czy SI próbuje formułować wnioski na temat własnych stanów wewnętrznych (ibidem: 443-446).

[21] Kwestię przysługiwania praw SI trafnie jak się zdaje podsumowują John-Stewart Gordon i Ausrine Pasvenskiene: „Jeśli oparta na krzemie istota – np. inteligentny, autonomiczny i świadomy siebie robot – spełni w jakimś momencie przyszłości odpowiednie kryteria (jakiekolwiek by one nie były), to ma pewne prawa bez względu na to, czy nam się to podoba, czy nie" (Gordon, Pasvenskiene 2021: 586). Precyzują oni jednocześnie, że „Bycie uprawnionym do praw człowieka wymaga [...] przede wszystkim ludzkich zdolności poznawczych, takich jak autonomia, inteligencja ogólna, racjonalność i na pewno jakaś forma samoświadomości" (ibidem 589).

SI jako moralny obiekt

W odniesieniu do uzasadnienia (2) warto wskazać, że w literaturze przedmiotu niektórzy (np. Darling 2016: 213-234) nawiązując do Kantowskich obowiązków niezupełnych[22], które mamy na przykład w stosunku do zwierząt wskazują, że analogiczny charakter mają ludzkie obowiązki wobec maszyn. Jak pisze Tomasz Pietrzykowski: „obowiązki niezupełne nie mają skonkretyzowanego beneficjenta, a tym samym nie towarzyszy im czyjekolwiek roszczenie o ich wykonanie" (Pietrzykowski 2015: 114). Kok-Chor Tan podkreśla, że obowiązki są niezupełne wtedy, gdy podmiot obowiązku, jego treść ani wobec kogo ma on zostać wypełniony nie zostały dokładnie sprecyzowane (Tan 2004: 50). Onora O'Neill określa ostatecznie obowiązki niezupełne jako nieskorelowane z uprawnieniami obowiązki moralne (O'Neill 2009: 428).

Wszystko to właściwie znaczy, że obowiązki niezupełne jednej strony relacji, będąc wymogami wyłącznie moralnymi, nie są skorelowane z uprawnieniami drugiej strony. W takim razie oparte na tych obowiązkach tak zwane „prawa maszyn" należałoby traktować jako fenomeny moralne. Powołując się na Maurice'a Cranstona nie nazwalibyśmy takich „praw" czymś, co „co jest obowiązkowe, co jest słuszne, co jest sprawiedliwe", a raczej czymś, co „miło byłoby zobaczyć, gdy się kiedyś wydarzy" (Cranston 1967: 53).

W tej sprawie zgodne z Kantowską charakterystyką obowiązków niezupełnych byłoby twierdzenie, że podmiot moralny w stosunku do istot czujących winien unikać niepotrzebnego przysparzania im cierpień. Spoczywanie na osobach wyłącznie obowiązków niezupełnych wobec tychże istot implikowałoby jednak ich status jako moralnych obiektów.

[22] Konkretne obowiązki niezupełne, o jakich pisze Kant, to na przykład nakaz rozwijania swoich talentów czy zakaz obojętności na to, że inni są w potrzebie. Różnicę między obowiązkami zupełnymi a niezupełnymi tak charakteryzuje Tomasz Pietrzykowski: „Kantowskie obowiązki zupełne to takie, których przedmiotem jest zachowanie należne konkretnemu uprawnionemu" (Pietrzykowski 2015: 114).

Istoty zdolne do odczuwania cierpienia winny być więc obiektem moralnych obowiązków spoczywających na osobach, ale nie jest to racja wystarczająca do tego, by owe istoty uznać za posiadaczy praw.

Co do uzasadnienia praw SI relacyjnie (3), to na przykład Peter-Paul Verbeek argumentuje, że „technologiczne artefakty nie są neutralnymi pośrednikami, ale aktywnie współkształtują ludzką egzystencję w świecie" (Verbeek 2011: 8). Twierdzi on nawet, że technologii da się przypisać intencjonalność (Verbeek 2005: 115). W jego przekonaniu „[...] podmiotowość moralna jest rozproszona zarówno pośród ludzi, jak i stykających się z nimi wytworów technologii (Verbeek 2008: 24). Martin Peterson i Andreas Spahn przekonują jednak, że niektóre argumenty Verbeeka zacierają fundamentalne rozróżnienie między rzeczywistością a naszym jej postrzeganiem, pomiędzy wpływem różnych bytów na siebie a ich integralnym związkiem oraz pomiędzy wpływem a intencjonalnością (Peterson, Spahn 2011: 415-417).

Wagę relacji między ludźmi i innymi bytami podkreśla też Mark Coeckelbergh. Uważa on, że stosunki te mają znaczenie moralne i nie można myśleć o moralności w oderwaniu od nich. Byty w tych relacjach (nazywanych przez niego ekologią społeczną) są wzajem zależne i dostosowują się do siebie (Coeckelbergh 2010: 217). Zdaniem badacza własności podmiotu, czy raczej przedmiotu, nie mają kluczowego znaczenia dla posiadania przezeń praw. Idzie w tym raczej o rolę, jaką przedmiot pełni w relacjach z człowiekiem.

Podejście relacyjne Coeckelbergha sugeruje, że nie powinniśmy zakładać moralnych własności jako jakoś przyczepionych do danego bytu. Zamiast tego waga moralna jest mu przyznawana w ramach dynamicznej relacji pomiędzy nim a ludźmi. W związku z tym Coeckelbergh proponuje uwzględnianie dwóch rodzajów praw: słabych (*soft*) i silnych (*hard*). Podczas gdy do tych drugich zaliczałyby się prawa człowieka, to pierwsze odnosiłyby się do niektórych robotów lub zwierząt i byłyby uzależnione od społecznej relacji pomiędzy tego rodzaju robotem (zwierzęciem) a człowiekiem (ibidem 218).

Wspomniane przez Coeckelbergha „słabe prawa" przywodzą na myśl krytyczne uwagi formułowane przez znanych etyków. Dla przykładu Amarty'a Sen wskazuje, że niektóre niekorelujące z obowiązkami zupełnymi problematyczne roszczenia, kandydujące do miana „praw", są tylko moralnymi postulatami czasem inspirującym legislację (Sen 2004" 318). Bernard Williams uważa, że są one wyrazem aspiracji w dążeniach do uzyskania pożądanych dóbr (Williams 2005: 64). O'Neill jest w tej kwestii bardziej stanowcza, bo projekty ustanawiania praw nieskorelowanych z obowiązkami, nazywa wręcz: „niedopuszczalnym oszustwem" (*unacceptable deception*) (O'Neill 2009: 428). Idzie o to, że konsekwencją traktowania praw jako postulatów jest to, że przestaną one być rzeczywistymi roszczeniami (ibidem: 435).

SI jako osoba i podmiot moralny

W rozważaniach o SI jako podmiocie moralnym trudno abstrahować od pojęcia „osoby" zwłaszcza, że źródłowo nie odnosiło się ono bynajmniej do bytu ludzkiego, a w każdym razie nie w pierwszym rzędzie. Kanoniczną definicję „osoby"[23] sprecyzował Boecjusz w księdze *O osobie i dwóch naturach*: „Osoba jest to poszczególna substancja natury rozumnej" (Boecjusz 2003: 259)[24]. Pomimo, że trudno o konsensus co do wyznaczników bycia osobą, to dość powszechna zgoda dotyczy tego, co się

[23] Łacińskie słowo „persona" pochodzi z greckiego „prosopon" oznaczającego maskę teatralną wyrażającą graną przez aktora postać.

[24] Odwołujemy się do tej definicji między innymi z dlatego, że odpowiada ona na problem związany ze skrajnymi przypadkami (*marginal cases*). Agrument ze skrajnych przypadków patrz (Coeckelbergh 2010: 212). Przywołanie „natury" w definicji Boecjusza wskazuje tu, że nie idzie tu o aktualny stan rzeczy, ale o pewną fundamentalną zdolność, potencjalność do bycia rozumnym. Z braku miejsca nie będziemy tu głębiej wchodzić w bardzo obszerną problematykę dotyczącą osoby. Była ona i jest podejmowana przez różne nurty filozofii. Zajmują się nią chociażby tomizm, personalizm, fenomenologia, filozofia dialogu, czy filozofia analityczna.

jej należy, a mianowicie: poszanowanie godności[25] i fundamentalnych praw. Tym, co w sposób szczególny zdaje się też wyróżniać osoby jest swoboda wyboru, wolność. Wolność w ujęciu Kantowskim ujawnia się w zdolności do sprzeciwu wobec empirycznych skłonności i pragnień. Polega ona na „[...] władzy [...] determinowania siebie [...] niezależnie od przyrodzonych instynktów" (Kant 2005: 82). Za Robertem Speamanem można powiedzieć, że najistotniejszą cechą osoby jest też możność wolnego odnoszenia się do swojej natury. Pisze on: „Od czego wolna jest osoba? Jest wolna od swojej własnej natury. Ma tę naturę, nie *jest* nią. Może się do niej odnosić w sposób wolny" (Spaemann 2001: 265).

Przechodząc do uzasadnienia praw SI nr (1) uznajemy, że osoba, bez względu na jej bardziej szczegółową charakterystykę, jest podmiotem moralnym. To jest takim, który nie tylko może dokonywać wyraźnych osądów moralnych, ale i jest na tyle kompetentny, aby je jakoś racjonalnie uzasadnić[26]. Przyjmujemy tu, że uznanie za osobę, niezależnie czy miałoby to dotyczyć bytu biologicznego, mechanicznego, cyfrowego, czy jeszcze innego, implikuje przysługiwanie takiemu bytowi praw analogicznych do praw człowieka.

Prezentację fundacjonistycznej koncepcji uzasadnienia praw SI jako autonomicznego bytu zaczniemy od rozważań wokół podmiotu moralnego. Przede wszystkim spróbujemy doprecyzować charakter bytu, który można by nazwać autonomicznym podmiotem moralnym. Dobrym tłem dla tego będzie krytyka interesujących skądinąd uwag Dane Gogoshin. Jest ona zdania, że nasze wzajemne działania moralne są programowane bardziej poprzez praktykę działania uwzględniającą moralną odpowiedzialność tak, aby zachowywać się zgodnie z zasadami

[25] Co do godności przysługującej osobie, to z ogromnej liczby dotyczących jej prac warto wskazać chociażby kilka pozycji: (Gogacz 1994: 113-140); (Piluś 1989: 168-69); (Szostek 1998: 42-63); (Mazurek 2001); (Sensen 2011: 71-91).

[26] Przychodzi tu na myśl stwierdzenie Arystotelesa, że szczególną własnością człowieka odróżniającą go od innych stworzeń jest zdolność rozróżniania dobra od zła i sprawiedliwości od niesprawiedliwości (Arystoteles 2005: 7).

i wartościami, niż poprzez działanie w imię racji moralnych. Te uwarunkowania skłaniają nas do przestrzegania norm moralnych na długo zanim będziemy w stanie pojąć moralne znaczenie naszych działań (Gogoshin 2021: 5-6). Nawiązując do Strawsona (2018) wskazuje ona, że jeśli moralność jest zestawem reguł, których przestrzeganie umożliwia istnienie społeczeństwa, a w efekcie wyższego dobra, to podmiotem moralnym (*moral agent*) jest ktoś, kto będzie przestrzegał tych reguł (ibidem: 7-8). Zaznacza ona też wyraźnie, że podmiot moralny może, ale wcale nie musi być podmiotem autonomicznym (czyli posiadającym wolną wolę w sensie Kantowskim).

Przyglądając się bliżej temu, co pisze Gogoshin zauważymy, że termin „odpowiedzialność moralna" (*moral responsibility*) oznacza w jej ujęciu tylko zdolność do moralnego postępowania. W takim rozumieniu byty są *moralnie odpowiedzialne* wówczas gdy zachowują się moralnie, czyli zgodnie z zasadami. W koncepcji Gogoshin status maszyn jako podmiotów moralnych opiera się więc na tym, że przestrzegają one norm. Co wydaje się kluczowe, że statusu tego nie deprecjonuje nawet ewentualna niezdolność do nie przestrzegania owych reguł. Jednakże w tym kontekście warto zauważyć, że bycie odpowiedzialnym kojarzy się z możnością nieprzestrzegania reguł i ponoszeniem za to konsekwencji. Moralność zakłada wolność, a więc podmiotem moralnym może być wyłącznie ten, kto może zarówno reguł przestrzegać, jak i nie przestrzegać. Problem ten trafnie ujął Kamil Mamak pisząc: „Tylko podmioty moralne (*moral agents*) mogą być odpowiedzialne za swoje działania, ponieważ rozumieją dobro i zło i są w stanie wybierać między nimi" (Mamak 2022: 3).

Nick Bostrom i Eliezer Yudkowsky zwracają uwagę, że w podobnych sytuacjach powinniśmy traktować sztuczny umysł w taki sam sposób, jak jakościowo identyczny z nim naturalny umysł ludzki (Bostrom, Yudkowsky 2011: 9). Bez względu na to, czy SI należałoby przyznać status moralnego podmiotu, czy tylko obiektu, to Bostrom i Yudkowsky zakładają, że twórcy lub właściciele systemu SI o statusie moralnym, mogą

mieć wobec niego, na podobieństwo stosunków pomiędzy rodzicami i dziećmi, specjalne obowiązki (ibidem).

W związku z tym zauważmy, że jeśli SI miałoby mieć prawa korespondujące z owymi obowiązkami twórców, to owa sztuczna inteligencja winna mieć zarazem specjalne obowiązki wobec swoich twórców. Odnosząc to do rozłożenia praw i obowiązków w relacjach pomiędzy dziećmi a rodzicami winniśmy doprecyzować, że zgodnie ze schematem Hohfelda rodzicielski obowiązek (*duty*) opieki nad dzieckiem skorelowany jest z prawem (*claim-right*) dziecka do tego, by się nim opiekowano. Z tym z kolei wiąże się rodzicielskie prawo do władzy (*power*) nad dzieckiem, z czym koreluje obowiązek dziecka do posłuszeństwa (*liability*) rodzicowi. Zauważmy jednak, że specyfika ludzkich relacji, w tym także rodzinnych, polega między innymi na tym, że ich uczestnicy są zdolni do niewypełniania spoczywających na nich obowiązków. Jakkolwiek takie przypadki w stosunkach między ludźmi zdają się czymś normalnym, to w analogicznej sytuacji ewentualne nieposłuszeństwo sztucznego superinteligentnego bytu musiałoby budzić nasz uzasadniony niepokój.

Mimo, że powstanie superinteligentnej maszyny wydaje się jedynie kwestią czasu, to powstanie sztucznej osoby może się okazać opcją wyłącznie teoretyczną. Komentując możliwość identyczności ludzkiego i sztucznego umysłu należałoby wskazać, że może to być trudne, między innymi z powodu specyfiki narodzin i funkcjonowania SI. Mowa tu na przykład o braku biologicznych uwarunkowań, braku właściwej człowiekowi inklinacji do życia we wspólnocie, być może nawet o braku tak zwanych uczuć wyższych itp. Zwróćmy w związku z tym uwagę, że SI trudno byłoby przypisać indywidualność, jej inteligencja jawi się ponadto jako nierelacyjna, podczas gdy ludzka inteligencja jest związana z

relacją z innymi osobami[27]. Ludzie nie są w stanie rozumieć świata inaczej niż w intencjonalnej relacji *Ja* nakierowanego na coś lub kogoś. Przy takich potencjalnych różnicach trudno *a priori* zakładać, by pomiędzy superinteligentną maszyną a człowiekiem mogła zachodzić jakościowa identyczność.

Podsumowując ten fragment rozważań należy podkreślić, że ostatecznie opowiadamy się za stanowiskiem, że tylko posiadanie własności analogicznych do tych przynależnych osobom mogłyby nadać SI status posiadacza praw. Znaczy to, że spośród możliwości wskazanych przez Gordona i Nyholma opowiadamy się za opcją (1). Odrzucamy (2) ponieważ obowiązki niezupełne (czyli moralne) nie implikują praw. Z kolei podejście relacjonalne (3) zdaje się zbyt pochopnie traktuje przedmiot relacji jako uprawniony jej podmiot.

Problemy dotyczące praw SI

David J. Gunkel w kilku swoich tekstach stawia problem: „czy roboty mogą i czy powinny mieć prawa?" (Gunkel 2018a: 87, podobnie idem 2018b: 2). Jego zdaniem „[...] już sama forma tego pytania ujawnia założenie o prawach jako rodzaju własności osobistej lub posiadania czegoś, co jednostka może mieć lub co powinno być jej nadane" (idem 2018a: 96). Zauważmy, że roboty wypełniając pewne kryteria, na przykład bycia osobą już je, niejako automatycznie, właśnie jako osoby uzyskują. W związku z tym pytanie: „czy powinny?" jest tu chyba nie na miejscu. Bardziej fundamentalnym pytaniem wydaje się: czy możemy i czy powinniśmy próbować stworzyć sztuczną osobę?[28]. I pochodne: czy kluczowe kryterium rozstrzygające o byciu istotą posiadającą prawa

[27] Na społeczne źródła i charakter ludzkiej inteligencji oraz na rolę jaką w tym wszystkim gra język zwraca uwagę (Railton 2020: 54-55).

[28] Sam Gunkel też zresztą w podobnym duchu przeformułowuje swoje pierwotne pytanie (Gunkel 2018a: 89).

(osobą) przypadkiem nie anihiluje, przynajmniej w części, a być może nawet w całości, owych praw? Idzie mianowicie o zdolność do nieposłuszeństwa, czyli do niewypełniania spoczywających na podmiocie obowiązków. Cieszenie się prawami jest bowiem uwarunkowane wypełnianiem obowiązków. Kto ich nie wypełnia, ten nie może cieszyć się pełnią praw. Tę prostą zależność dostrzeżemy w ramach aktów retrybucji. Zauważmy bowiem, że przestępstwo, czyli wystąpienie przeciw prawom innego podmiotu skutkuje ograniczeniem jakichś praw przestępcy. Zatem kluczowe pytanie nie brzmi: czy roboty mogą i czy powinny mieć prawa?, ale: czy roboty mogą i czy powinny być osobami? Naturalnie negatywna odpowiedź na pierwszą część tego pytania unieważnia całą kwestię.

Innym problemem jest zakres praw dostępnych sztucznym osobom. Jeśli nawet okażą się one podmiotami moralnymi, to nie jest jeszcze powiedziane, że muszą im przysługiwać wszystkie prawa, w tym samym stopniu, co ludziom. Mathew Liao słusznie bowiem argumentuje, że sztuczne osoby mogą mieć ograniczenia w pewnych sferach. Podaje on przykład prawa do reprodukcji. Ze względu na ograniczenia dostępnych zasobów i potencjalną gigantyczną skalę reprodukcji sztucznych osób, tego rodzaju prawo mogłoby być dla sztucznych osób limitowane (Liao 2020: 495).

Kolejny wart omówienia problem dotyczy wątpliwości co do statusu maszyn. Zajmując się problematyką sztucznej osoby należałoby bowiem ostatecznie zapytać: co w praktyce oznaczałoby jej powstanie? Jeśli ultrainteligentna maszyna okaże się sztuczną osobą, to jako takiej należą jej się prawa przysługujące osobom, choćby nawet nie była ona człowiekiem. W takim razie nie można być jej właścicielem ani zmuszać jej do pracy, gdyż wypełniałoby to znamiona niewolnictwa w jego ścisłym sensie. Taki zakaz poza wszystkim innym zabezpieczałby, w jakiejś mierze, przed możliwością ustanowienia bardzo niebezpiecznego monopolu na korzystanie z tego rodzaju ultrainteligentnej maszyny.

W związku z tą kwestią Joanna Bryson uważa na przykład, że roboty powinny pozostawać w służbie ludzi jako narzędzia (niewolnicy) zaspokajające ich potrzeby i zaprojektowane jako twory nie będące moralnymi podmiotami (Bryson 2010: 63-74). Twierdzi ona, że „Błędem byłoby pozwolić ludziom myśleć, że ich roboty są osobami" między innymi dlatego, że prowadziłoby to do nietrafnego rozłożenia odpowiedzialności za ich działania (ibidem: 65). Jej zdaniem przypisujemy robotom ludzkie cechy po części z powodu naszej niepewności co to właściwie znaczy być człowiekiem (ibidem: 66)[29]. Ostatecznie Bryson uważa, że mamy wręcz obowiązek nie konstruowania takich robotów, wobec których mielibyśmy obowiązki (Bryson 2018: 15).

Jej podejście do tego problemu wywołało żywą reakcję. Podniosła się na przykład krytyka dotycząca niewolniczego statusu maszyn, w tym zwłaszcza wobec dehumanizacji, do której prowadzi niewolnictwo. Pojawiły się jednak także głosy wspierające twierdzenia Bryson. Na przykład Abeba Birhane i Jelle van Dijk zwracają uwagę, że „dehumanizacja" robotów jest niemożliwa, bo nie można odczłowieczyć czegoś, co nigdy nie było człowiekiem (Birhane, van Dijk 2020: 207).

W istocie postawienie zarzutu dehumanizacji w takim kontekście implikuje, że stawiający go sam niejako „humanizuje" roboty. Birhane i van Dijk odnosząc się do tekstu Bryson piszą: „Naszym punktem wyjścia nie jest odmawianie robotom «praw», ale zaprzeczanie temu, że roboty są istotami, którym można przyznać lub odmówić praw. Sugerujemy, że wyobrażanie sobie robotów jako niewolników nie ma sensu, ponieważ «niewolnik» należy do kategorii istot, którymi roboty nie są". W związku z tym uważają oni, że „Argumentowanie za prawami robotów na podstawie przyszłych wizji czujących maszyn jest w najlepszym razie czysto teoretycznym spekulowaniem" (ibidem: 208). Autorzy ci są

[29] Warto odotować, że Jacob Turner wspomina w tym kontekście o „sofizmacie androida" (*Android Fallacy)*. Jest to błąd polegający na zjednaniu najszerzej pojmowanej „osobowości" z „człowieczeństwem" i z prawami przynależnymi do tego ostatniego (Turner 2019: 189).

bardzo sceptyczni co do rewolucyjnej jakościowej zmiany jaką przyniosą bardziej rozwinięte SI, piszą oni: „Wygląda na to, że «AGI», «technologiczna osobliwość», czy «super-inteligencja» są dla techno-optymistów tym, czym koniec świata jest dla kultów religijnych" (ibidem: 210).

Inną, choć też krytyczną wobec praw robotów, optykę prezentuje Lantz Fleming Miller, według niego istniejąca pomiędzy robotami a ludźmi różnica ontologiczna uzasadnia przypisanie im odmiennych praw (Miller 2015: 380). W przeciwieństwie do celowo przez kogoś skonstruowanych maszyn ludzie nie mają wyznaczonego z zewnątrz celu swojego istnienia, którą to właściwość człowieka nazywa on „egzystencjalną neutralnością normatywną" (ibidem: 383). Ową ontologiczną różnicę w statusie moralnym Miller uważa za wystarczającą podstawę dla odmowy przyznania automatom praw człowieka (ibidem: 379). Komentujący to Kamil Mamak stwierdza: „Roboty nie będą miały takich samych jak przysługujące ludziom legalnych praw i obowiązków ponieważ treść praw człowieka jest powiązana z jego biologią, której roboty nie mają. Innymi słowy, ze względu na różnice ontologiczne tych bytów, «prawa robota» nigdy nie będą w pełni pokrywać się z «prawami człowieka»" (Mamak 2022: 6).

Gdyby kiedyś SI okazała się osobą, to byłaby praktycznie nieśmiertelną istotą o trudnych dziś do wyobrażenia możliwościach. W związku z tym zdaniem Erica Schwitzgebela i Mary Garzy rozsądny środek ostrożności wobec SI polegałby na polityce projektowej wyłączonego środka (*Excluded Middle*). Miałaby ona unikać tworzenia takich SI, co do których *byłyby wątpliwości*, czy należą im się takie same względy, co ludziom (Schwitzgebel, Garza 2020: 466)[30].

[30] Postulują oni też nienaruszanie praw SI ufundowanych na połączonych kryteriach utylitarystycznych (maksymalizacja szczęścia i unikanie cierpienia) i deontologicznych (respekt dla racjonalnej istoty) (Schwitzgebel, Garza 2020: 462).

Powyższa propozycja zastosowania zasady wyłączonego środka, jakkolwiek wydaje się rozsądna, to jednak nie rozwiązuje głównego problemu. Wprawdzie radzi sobie z sytuacjami niejasnymi[31], ale ważniejsze jest to, co mielibyśmy począć w sprawie bytu, który jednak należałoby określić jako sztuczną osobę.

W tej zaś dziedzinie możliwe wydają się przynajmniej dwa rozwiązania dające się dość łatwo zastosować w praktyce. Pierwsze sprowadzałoby się do zakazu tworzenia wysoko rozwiniętych inteligentnych maszyn bliskich statusowi sztucznej osoby. Drugie polegałoby na tym, by zamiast stawiać na rozwój SI w postaci reprezentowanej przez ASI lub AGI, postawić na doskonalenie IA (*Intelligence Augmentation* lub *Intelligence Amplification*, na jej temat zob. Biocca 1996: 59-75), czyli wsparcia człowieka elementami wysoko zaawansowanych urządzeń[32] z zastrzeżeniem powszechnej dostępności tego rodzaju wspomożenia.

Do zalet tego rozwiązania należy to, że udoskonalony cyfrowo człowiek, byłby tylko jednym z mnóstwa innych udoskonalonych cyfrowo ludzkich osób, nad którymi nie będzie miał żadnej istotnej przewagi. Ponadto nie będzie żadnych wątpliwości co do osobowego statusu wszystkich takich podmiotów. Będzie również wiadomo jak je karać za niewykonywanie swoich obowiązków w postaci nienaruszania praw innych osób. Związane z tym, odnoszące się do ludzkich osób reguły, normy, procedury są już bowiem od dawna dopracowane. Tego rodzaju „hybrydy" nie będą tworzyły żadnego bytu o jakiejś nowej szczególnej

[31] W tej z kolei sprawie Schneider proponuje, by normalną częścią badań nad rozwiniętymi systemami SI były testy na świadomość (Schneider 2020: 454).

[32] IA „ma na celu pracę z ludźmi i skupienie się na budowaniu systemów, które zwiększają i wspierają ludzkie poznanie" (Hassani, Silva, Unger, Mazinani, Mac Feely 2020: 147). Takie wzmocnienie człowieka elementami sztucznej inteligencji mogłoby się dokonać na przykład poprzez głębokie implanty w tkankę mózgu. Główna różnica pomiędzy taką hybrydą a ultrainteligentną maszyną polega na tym, że centrum całego systemu decyzyjnego IA jest człowiek, podczas gdy w przypadku SI jest nim technologia. W gruncie rzeczy SI miałaby być autonomiczną sztuczną świadomością (umysłem) w maszynie, podczas gdy hybryda wykorzystująca IA byłaby połączeniem ludzkiej świadomości (umysłu) z podporządkowaną mu maszyną (ibidem: 148).

specyfice podobnie, jak nie tworzy go człowiek korzystający z laptopa czy latający samolotem.

Kolejny problem wiąże się ewentualną moralnością maszyn. Gdy przyjrzymy się naszym własnym etycznym imponderabiliom, to okazuje się, że część z nich w żaden sposób nie przystaje do funkcjonowania maszyn, które nie mają życia rodzinnego, przyjaciół, osób ukochanych, społeczności, czy instytucji, z którymi mogłyby się identyfikować, do których przynależeć.

Ustalenie jakichś stałych moralnych punktów odniesienia wydaje się jeszcze bardziej skomplikowane gdy uświadomimy sobie, że kwestia „jaką etyką winna kierować się sztuczna osoba?" może być równie daleka od konsensusu jak w przypadku pytania: „jaką etyką winni kierować się ludzie?"[33]. Dodatkową trudność w tym względzie sprawia to, ze sztuczna osoba znacząco by się od nas różniła. Nie miałaby wielu uczuć, które nam wydają się oczywiste. Być może nie miałaby też preferencji nie tylko co do konkretnych osób, ale także wobec przyszłości, czy teraźniejszości (jej preferencja czasowa mogłaby być bliska zeru), piękna, czy swojskości.

Właściwie porządek normatywny sztucznej osoby nie opierałby się chyba na moralności, zamiast tego mógłby jednak odwoływać się do któregoś z uniwersalistycznych nurtów etyki. Przykładowo, niska preferencja czasowa przybliżałaby modus działania takiej SI do zaleceń formułowanych w ramach etyki cnoty. Z drugiej strony bardzo prawdopodobne nakierowanie SI na maksymalizację użyteczności oraz traktowanie każdego tak samo („za jednego") mogłoby być impulsem dla ukazania jej specyfiki w perspektywie utylitaryzmu. Jej czysto racjonalne podejście tworzyłoby zaś jakąś koneksję z deontologią. W tym kontekście problemem jest jednak to, że ani wartości, ani etyka (a już tym bardziej

[33] Jak zwraca na to uwagę Christopher Barr, „jeśli nie znamy zasad, według których dokonujemy etycznych rozstrzygnięć, to trudno zakładać, że stworzymy AGI z analogiczną do naszej psychologią moralną" (Barr 2017: 28).

moralność) nie są cyfrowymi danymi i trudno powiedzieć, czy dadzą się do nich sprowadzić. Jak się wydaje, z racji tej odmienności oraz związanej z nią nieprzewidywalnych cech sztucznej osoby pojawiają się obawy przed powstaniem ultrainteligentnych maszyn[34].

Kolejny problem, związany już nie tyle z prawami, co z odpowiedzialnością robotów znany jest jako „luka odpowiedzialności" (*responsibility gap*). Polega ona na tym, że roboty wykonują coraz więcej zadań zastępując nie tylko pojedynczych ludzi, ale całe firmy. W związku z tym zwiększa się prawdopodobieństwo popełnienia błędu, spowodowania szkody, zrobienia czegoś o nieprzewidzianych konsekwencjach, za które żadna osoba nie ponosi odpowiedzialności. Dalszą postacią powyższego problemu jest tak zwana luka odwetowa (*retribution gap*). Powstaje ona, gdy ludzie domagają się kary dla sprawcy szkody, ale nie znajdują odpowiedniego podmiotu dla zaspokojenia tego pragnienia" (Kraaijeveld 2020: 1317)[35].

Przedostatni problem, który omówimy dotyczy skutków błędnej oceny. Prawne rozwiązania w sferze uprawnień i obowiązków maszyn są pochodną rozstrzygnięć co do specyfiki tego bytu w wymiarze ontologicznym i etycznym Jakie są konsekwencje błędu w tej materii? Zasadniczo dwie: 1. przyznanie praw człowieka bytom nie będącym osobami; 2. nieprzyznanie praw człowieka sztucznym osobom. Pierwsza opcja, która wielu wydaje się lepsza niż druga, skutkuje nałożeniem na ludzi obowiązków, które sprawiedliwie rzecz biorąc są nieuzasadnionymi ciężarami. To zaś w konsekwencji oznacza na ogół ograniczenie rzeczywistych praw przynależnych osobom ludzkim. Co więcej, przyznanie

[34] Klasyczny już dziś tekst o zagrożeniach dla funkcjonowania ludzkości, uwzględniających także te płynące ze strony SI, patrz (Bostrom 2002: 1-31). Nick Bostrom owe zagrożenia definiuje następująco: „Są to zagrożenia, które mogą spowodować nasze wyginięcie lub zniszczyć potencjał pochodzącego z Ziemi inteligentnego życia" (ibidem: 1).

[35] Odnosząc się do tej kwestii Colin Jones pisze o modelu, w którym odpowiedzialność za działania robotów mogłaby mieć charakter analogiczny do odpowiedzialności rodziców za działania ich dzieci. I w tym sensie roboty byłyby czymś w rodzaju „wiecznych dzieci" (*perpetual children*) (Jones 2019: 410).

takich praw będzie zarazem narzuceniem obowiązków zapewnienia takim maszynom podstawowych dóbr, uczestnictwa w życiu politycznym, nadanie osobowości prawnej[36], zabezpieczenia ich własności, dopuszczenie ich do konkurowania z ludźmi o ograniczone zasoby itd. Wszystko to może prowadzić do resentymentu wobec maszyn, a być może także do prób ich niszczenia.

Na koniec pozostawiliśmy najbardziej fundamentalny problem. Otóż zauważmy, że ze względu na nasze ograniczenia poznawcze nie jesteśmy w stanie potwierdzić z zewnątrz istnienia pewnych stanów mentalnych innych podmiotów[37]. Jakkolwiek w przypadku innych ludzi pomijamy tę fundamentalną niepewność zakładając analogię z naszymi własnymi stanami mentalnymi, to w przypadku maszyn taka opcja nie wchodziłaby w grę[38].

Zakończenie

W artykule, z konieczności jedynie pobieżnie zasygnowaliśmy pewne kwestie wiążące się z rozwojem SI, na które najprawdopodobniej, raczej wcześniej niż później, będzie trzeba poszukać odpowiedzi. W związku z nimi odwołaliśmy się do charakterystyki sztucznej inteligencji, praw podmiotowych obiektów i podmiotów moralnych, a także specyfiki sztucznej osoby oraz uzasadnienia jej ewentualnych praw. W tej dziedzinie powołując się na Gordona i Nyholma wskazaliśmy trzy możliwości. Pierwsza odwoływała się do charakterystyki sztucznej inteligencji jako podmiotu moralnego (sztucznej osoby). Druga wywodziła prawa

[36] Na temat osobowości prawnej sztucznej inteligencji zob. też (Biczysko-Pudełko, Szostek 2019).

[37] Z tego powodu w odniesieniu do maszyn niektórzy teoretycy zajmują stanowisko moralnego behawioryzmu, to znaczy uznają, że na podstawie zewnętrznych objawów, jakie dadzą się zaobserwować u pewnych podmiotów, winniśmy uznawać istnienie ich wewnętrznych stanów (zob. Danaher 2020: 2023-2049).

[38] W sposób oczywisty sytuacja ta przywodzi na myśl słynny tekst Thomasa Nagela włącznie z jego konkluzjami: *What Is It Like to Be a Bat?* (Nagel 1974).

sztucznej inteligencji ze spoczywających na ludziach obowiązków niezupełnych wobec niej. Trzecia fundowała prawa na moralnych relacjach między ludźmi i maszynami. W tekście przytaczaliśmy argumentację wielu teoretyków dotyczącą wszystkich tych opcji. Słabością drugiej z nich było przede wszystkim to, że ludzkie obowiązki niezupełne nie implikują praw SI. W ramach trzeciej ewentualności potencjalne prawa SI bardziej przypominały moralne postulaty bądź aspiracje niźli rzeczywiste roszczenia. Z przedstawionych możliwości pierwsza, w której SI spełniałaby kryteria moralnego podmiotu (osoby) wydaje się najbliższa koncepcji SI jako posiadacza praw. Jednakże nawet ta ewentualność rodzi wiele wątpliwości i problemów wskazywanych przez licznych badaczy, a których kilka staraliśmy się naszkicować.

Podsumowując można stwierdzić, że problematyka ta musi radzić sobie z wieloma niewiadomymi. Przypisanie praw maszynom, nawet tym najbardziej zbliżonym do człowieka wydaje się zderzać z rozległą i całkiem dobrze uzasadnioną krytyką.

Literatura

Arystoteles (2005). *Polityka*. Wrocław: Ossolineum.

Barr., Ch. (2017). Building Morality: A New Strategy for Creating Human-Like Moral Psychology in Artificial General Intelligence. *Lawrence University Honors Projects* 112, 1-37.

Beitz, Ch.R. (2009). *The Idea of Human Rights*. Oxford: Oxford University Press.

Biczysko-Pudełko, K., Szostek, D. (2019). Koncepcje dotyczące osobowości prawnej robotów – zagadnienia wybrane. *Prawo Mediów Elektronicznych* 2, 9-15.

Biocca, F. (1996). Intelligence augmentation: The vision inside virtual reality. *Advances in Psychology* 113, 59-75.

Birhane, A., van Dijk, J. (2020). Robot rights? Let's talk about human welfare instead. [w:] *AIES'20: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, February 7–8, 2020, New York*. New York, https://doi.org/10.1145/3375627.3375855.

Boecjusz (2003). *O pociechach filozofii. Traktaty teologiczne*. Warszawa: DeAgostini.

Bostrom, N., Yudkowsky, N. (2011). *The Ethics of Artificial Intelligence*. file:///C:/Users/UWM/Downloads/artificial-intelligence.pdf.

Bostrom, N. (2002). Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards. *Journal of Evolution and Technology* 9(1), 1-31.

Bryson, J.J. (2018). Patiency is not a virtue: The design of intelligent systems and systems of ethics. *Ethics and Information Technology* 20, 15-26.

Bryson, J.J. (2010). Robots should be slaves. [w:] Wilks, Y. (Ed.), *Close Engagements with Artificial Companions: Key social, psychological, ethical and design issues*. Amsterdam: John Benjamins, 63-74.

Buchanan, A. (2010). The Egalitarianism of Human Rights. *Ethics* 120, 679-710.

Coeckelbergh, M. (2012). *Growing Moral Relations. Critique of Moral Status Ascription*. London: Palgrave Macmillan.

Coeckelbergh, M. (2010). Robot rights? Towards a social-relational justification of moral consideration. *Ethics and Information Technology* 12 (3), 209-221.

Copeland, J. (2000). What is Artificial Intelligence?. *AlanTuring.net.*, http://www.alanturing.net/turing_archive/pages/Reference%20Articles/What%20is%20AI.html.

Cranston, M. (1983). Are There Any Human Rights. *Daedalus* 112(4), 1-17.

Cranston, M. (1967). Human Rights-Real and Supposed. [w:] Raphael, D.D. (Ed.), *Political Theory and the Rights of Man*. Bloomington, IN: Indiana University Press, 43-53.

Danaher, J. (2020). Welcoming Robots into the Moral Circle: A Defence of Ethical Behaviourism. *Science and Engineering Ethics* 26, 2023-2049.

Darling, K. (2016). Extending Legal Protection to Social Robots: The Effects of Anthropomorphism, Empathy, and Violent Behavior towards Robotic Objects. [w:] Calo, R., Froomkin, A.M., Kerr, I. (Eds.), *Robot Law*. Cheltenham: Edward Elgar Publishing, 213-234.

Dumouchel, P. (2019). Intelligence, Artificial and Otherwise. *Forum Philosophicum* 24(2), 241-258.

Floridi, L. (2013). *The ethics of information*. Oxford: Oxford University Press.

Frankena, W.K. (1973). *Ethics*. Englewood Cliffs: Prentice-Hall.

Gogacz, M. (1994). Filozoficzna identyfikacja godności osoby. [w:] Czerkawski, J. (red.), *Zagadnienia godności człowieka*. Lublin: Wydawnictwo KUL, 113-140.

Gogoshin, D.L. (2021). Robot Responsibility and Moral Community. *Frontiers in Robotics and AI* 8,. https://www.frontiersin.org/articles/10.3389/frobt.2021.768092/full, 1-13.

Good, I.J. (1965). Speculations Concerning the First Ultraintelligent Machine. *Advances in Computers* 6, 31-88.

Gordon, J.S., & Nyholm, S. (2021). Ethics of Artificial Intelligence. W: *Internet Encyclopedia of Philosophy*. https://iep.utm.edu/ethics-of-artificial-intelligence/#SH2g.

Gordon, J.S., Pasvenskiene, A. (2021). Human Rights for Robots? A Literature Review. *AI and Ethics* 1,. https://doi.org/10.1007/s43681-021-00050-7, 597-591.

Gunkel, D.J. (2018a). The other question: can and should robots have rights?. *Ethics and Information Technology* 20, 87-99.

Gunkel, D.J. (2018b). *Robot Rights*. London: The MIT Press.

Harel, A. (2005). Theories of Rights. [w:] Golding, M.P., Edmundson, W.A. (Eds.), *The Blackwell Guide to the Philosophy of Law and Legal Theory*. Malden: Blackwell, 191-206.

Hassani, H., Silva, E.S., Unger, S., Mazinani, M.T., Mac Feely, S. (2020). Artificial Intelligence (AI) or Intelligence Augmentation (IA): What Is the Future? *AI* 1, 143-155.

Hohfeld, W.N. (1917). Fundamental Legal Conceptions as Applied in Judicial Reasoning. *Faculty Scholarship Series*. Paper 4378, http://digitalcommons.law.yale.edu/fss_papers/4378.

Janowska, M. (2015). Podmiotowość prawna sztucznej inteligencji? [w:] Bielska-Brodziak, A. (red.), *O czym mówią prawnicy, mówiąc o podmiotowości*. Katowice: Wydawnictwo Uniwersytetu Śląskiego.

Jones, C.P.A. (2019). The Robot Koseki: A Japanese Law Model for Regulating Autonomous Machines. *Journal of Business & Technology Law* 14/2, 403-467.

Kant, I. (2005). *Ugruntowanie metafizyki moralności*. Kraków: Zielona Sowa.

Kraaijeveld, S.R. (2020). Debunking (the) Retribution (Gap). *Science and Engineering Ethics* 26(3), https://doi.org/10.1007/s11948-019-00148-6, 1315-1328.

Kramer, M., Simmonds, N.E., Steiner, H. (1998). *A Debate Over Rights: Philosophical Enquiries*. Oxford: Clarendon Press.

Legg, S., Hutter, M. (2007). A collection of definitions of intelligence. [w:] Goertzel, B., Wang, P. (Eds.), *Artificial General Intelligence: Concepts, Architectures and Algorithm*. Amsterdam: IOS Press, 17-24

Liao, S.M. (2020). The Moral Status and Rights of Artificial Intelligence. [w:] Liao, S.M. (Ed.), *Ethics of Artificial Intelligence*. Oxford: Oxford University Press, 480-503.

Mamak, K. (2022). Humans, Neanderthals, robots and rights. *Ethics and Information Technology* 24(3), https://link.springer.com/content/pdf/10.1007/s10676-022-09644-z.pdf, 1-9.

Marshall, T.H. (1950). *Citizenship and Social Class*. Cambridge: Cambridge University Press.

Mazurek, J. (2001). *Godność osoby ludzkiej podstawą praw człowieka*. Lublin: Wydawnictwo KUL.

Miller, L.F. (2015). Granting Automata Human Rights: Challenge to a Basis of Full-Rights Privilege. *Human Rights Review* 16, 369-391.

Moor, J.H. (2006). The Nature, Importance, and Difficulty of Machine Ethics. *IEEE Intelligent Systems* 21(4), 18-21.

Nagel, T. (1974). What Is It Like to Be a Bat? *The Philosophical Review* 83(4), 435-450.

O'Neill, O. (2009). The Dark Side of Human Rights. [w:] Christiano, T., Christman, J. (Eds.), *Contemporary Debates in Political Philosophy*. Oxford: Blackwell, 423-436.

Peters, T. (2019). Artificial Intelligence versus Agape Love Spirituality in a Posthuman Age. *Forum Philosophicum* 24(2), 259-278.

Peterson, M., Spahn, A. (2011). Can Technological Artefacts Be Moral Agents?. *Science and Engineering Ethics* 17, 411-424.

Pietrzykowski, T. (2015). Kant, Korsgaard i podmiotowość moralna zwierząt. *Archiwum Filozofii Prawa i Filozofii Społecznej* 2, 106-119.

Piluś, H. (1989). O godności człowieka jako osoby. *Studia filozoficzne* 7-8, 163-182.

Railton, P. (2020). Ethical Learning, Natural and Artificial. [w:] Liao, S.M. (Ed.), *Ethics of Artificial Intelligence*. Oxford: Oxford University Press, 45-78.

Russell, S., Norvig, P. (2003). *Artificial Intelligence: A Modern Approach*. Upper Saddle River: Prentice Hall.

Schneider, S. (2020). How to Catch an AI Zombie: Testing for Consciousness in Machines. [w:] Liao, S.M. (Ed.), *Ethics of Artificial Intelligence*. Oxford: Oxford University Press, 439-458.

Schwitzgebel, E., Garza, M. (2020). Designing AI with Rights, Consciousness, Self-Respect, and Freedom. [w:] Liao, S.M. (Ed.), *Ethics of Artificial Intelligence*. Oxford: Oxford University Press, 459-479

Searle, J. (2004). *Mind: a brief introduction*. Oxford: Oxford University Press.

Sen, A. (2004). Elements of a Theory of Human Rights. *Philosophy and Public Affairs* 32(4), 315-356.

Sensen, O. (2011). Human dignity in historical perspective: The contemporary and traditional paradigms. *European Journal of Political Theory* 10(1), 71-91.

Shue, H. (1996). *Basic Rights*. Princeton: Princeton University Press.

Spaemann, R. (2001). *Osoby. O różnicy między czymś a kimś*. Warszawa: Oficyna Naukowa.

Sparrow, R. (2004). The Turing Triage Test. *Ethics and Information Technology* 6(4), 203-213.

Strauss, L. (1953). *Natural Right and History*. Chicago: University of Chicago Press.

Strawson, P.F. (1980). *Indywidua. Próba metafizyki opisowej*. Warszawa: PAX.

Szostek, A. (1998). Rola pojęcia godności w etyce. [w:] Szostek, A. *Wokół godności, prawdy i miłości. Rozważania etyczne*. Lublin: Wydawnictwo KUL, 42-63.

Sztuczna inteligencja dla Europy (2018). Komunikat Komisji do Parlamentu Europejskiego, Rady Europejskiej, Rady, Europejskiego Komitetu Ekonomicznospołecznego i Komitetu Regionów, COM (2018) 237 final, https://eur-lex.europa.eu/legal-content/PL/TXT/?uri=COM%3A2018%3A237%3AFIN.

Taddeo, M., Floridi, L. (2018). How AI can be a force for good. *Science* 361 (6404), 751-752.

Tan, K.Ch. (2004). *Justice Without Borders: Cosmopolitanism, Nationalism and Patriotism*. Cambridge: Cambridge University Press.

Tasioulas, J. (2012). On the Nature of Human Rights. [w:] Ernst, G., Heilinger, J.C. (Eds.), *The Philosophy of Human Rights: Contemporary Controversies*. Berlin: Walter de Gruyter, 17-60.

Tierney, B. (1997). *The Idea of Natural Rights: Studies on Natural Rights, Natural Law, and Church Law 1150–1625*. Cambridge: Eerdmans Publishing.

Tierney, B. (2004). The Idea of Natural Rights-Origins and Persistence. *Northwestern Journal of International Human Rights* 2(1), 1-12.

Tuck, R. (1979). *Natural Rights Theories: Their Origin and Development*. Cambridge: Cambridge University Press.

Turner, J. (2019). *Robot rules. Regulating artificial intelligence*. Cham: Palgrave Macmillan.

Verbeek, P.P. (2011). *Moralizing Technology: Understanding and Designing the Morality of Things*. Chicago: University of Chicago Press.

Verbeek, P.P. (2008). Obstetric ultrasound and the technological mediation of morality: A post phenomenological analysis. *Human Studies* 31, 11-26.

Verbeek, P.P. (2005), *What things do: Philosophical reflections on technology, agency and design*. University Park, PA: Penn State University Press.

Villey, M. (1983). *Le Droit et les droits de l'homme*. Paris: PUF.

Waldron, J. (1989). Rights in Conflict. *Ethics* 99 (3), 503-519

Wenar, L. (2005). The nature of rights. *Philosophy and Public Affairs* 33 (3), 223-252.

Williams, B. (2005). Human Rights and Relativism. [w:] Hawthorn, G. (Ed.), *In the Beginning was the Deed: Realism and Moralism in Political Argument*. Princeton: Priceton University Press, 62-74.

Zalewski, T. (2020). Definicja sztucznej inteligencji. [w:] Lai, L., Świerczyński, M. (red.), *Prawo sztucznej inteligencji*. Warszawa: C.H. Beck, 1-14.