O moralnej odpowiedzialności HAL-a 9000[1], czyli etyka sztucznej inteligencji w praktyce. Czy potrzebujemy definicji sztucznej inteligencji?

dr *Michał Nowakowski*[2]

HAL: [His shutdown] *I'm afraid. I'm afraid, Dave. Dave,* ***my mind is going****...*
2001: Odyseja kosmiczna

W ostatnim czasie mamy do czynienia z rozwojem różnych metod z obszaru tzw. sztucznej inteligencji (SI), w tym brane są pod uwagę aspekty etyczne w kontekście nowych technologii. Powoduje to, przynajmniej na dziś, niekoniecznie efektywne dyskusje i rozważania, których wartość praktyczna w kontekście rozwoju SI o obecnym poziomie technologicznym wydaje się niewielka. Jednocześnie na plan pierwszy w kontekście szeroko rozumianej etycznej sztucznej inteligencji wysuwa się nie tyle problematyka wyboru „metody" implementacji norm etycznych, ile samego określenia, względem „kogo" lub „czego" stosujemy dane wymogi. Z tego powodu niniejszy artykuł poświęcony jest głównie zagadnieniu pojęcia (systemów) sztucznej inteligencji, które stanowić powinno punkt wyjścia dla dyskusji nad tzw. moralnością maszyn.

Uwagi wstępne

W ostatnich latach rozwój różnych metod z obszaru tzw. SI[3] spowodował, że na nowo rozgorzała debata nad wieloma aspektami etycznymi[4] w kontekście nowych technologii, choć zagadnienie to jest obecne w dyskusjach akademickich od wielu lat[5]. Projekt rozporządzenia ws. sztucznej inteligencji, który ma ustanowić ramy prawne dla zastosowania SI czy opracowywane wytyczne UNESCO w sprawie etycznej sztucznej inteligencji[6], niewątpliwie stanowią próbę zaadresowania wyzwań związanych z postępującą algorytmizacją życia i jej wpływem na człowieka[7]. W doktrynie pojawiają się jednak liczne wątpliwości co do tego, czy (i jeżeli tak, to jak) możliwe jest przeniesienie na grunt (nie)rozumnych maszyn pewnych norm etycznych[8]. Jest to w dużej mierze konsekwencja debaty nad przyszłością sztucznej inteligencji[9] i możliwości stworzenia tzw. ogólnej sztucznej inteligencji[10] (*Artificial General Intelligence*). Po-

[1] Fikcyjny komputer przedstawiony w książce 2001: Odyseja kosmiczna autorstwa *A.C. Clarke'a* i w filmie o tym samym tytule.

[2] Head of NewTech w NGL Advisory, współpracownik w Śląskim Centrum Inżynierii Prawa, Technologii i Kompetencji Cyfrowych – CyberScience. ORCID: 0000-0002-8841-6566.

[3] Problematyka definicji sztucznej inteligencji stanowi wyzwanie dla wielu naukowców oraz prawodawców. Próby wprowadzenia takiego pojęcia pojawiają się obecnie na poziomie UE, choć liczne definicje można też spotkać w wielu opracowaniach organizacji międzynarodowych, np. OECD, które definiuje sztuczną inteligencję jako „system maszynowy, zdolny do wpływania na środowisko poprzez wytwarzanie «rezultatów» [*output*] (predykcji, rekomendacji czy decyzji, dla określonego zestawu celów). Taki system wykorzystuje dane (w tym wejściowe) wytwarzane przez człowieka i/lub maszyny do postrzegania realnych lub wirtualnych środowisk; przetwarzania obserwacji na modele poprzez analizę z użyciem automatycznych rozwiązań lub manualnie oraz wykorzystanie modeli do formułowania opcji lub wyników. Systemy sztucznej inteligencji mają różny poziomy autonomiczności" – https://oecd.ai/en/ai-principles (dostęp z 25.10.2021 r.). Zagadnieniu temu przyjrzymy się w dalszej części opracowania.

[4] *V.C. Muller* wskazuje na pojęcie etyki maszyn, które rozumieć należy jako etykę dla maszyn (jakkolwiek je zdefiniujemy – przyp. *M.N.*), w ramach której maszyny traktujemy nie jako obiekty, ale podmioty. *V.C. Müller* (forthcoming 2021), Ethics of artificial intelligence, [w:] *A. Elliott* (red.), The Routledge social science handbook of AI (London: Routledge), s. 14. Wydaje się jednak, że bardziej właściwe jest przyjęcie mniej „człowieczej" definicji etyki sztucznej inteligencji, w tym sensie, że etyka pozostaje etyką człowieka, a modele sztucznej inteligencji są jedynie narzędziem w ręku człowieka i w tym sensie realizują jego moralność. Zagadnienie to objaśnione zostanie w dalszej części opracowania.

[5] Zob. *S.M. Liao* (red.), Ethics of Artificial Intelligence, Oxford 2020, *passim*.

[6] UNESCO, First draft of the recommendation on the ethics of artificial intelligence, SHS/BIO/AHEG-AI/2020/4 REV.2, https://unesdoc.unesco.org/ark:/48223/pf0000373434 (dostęp z 26.10.2021 r.).

[7] Bardzo rozbudowane badania nt. wpływu uczenia maszynowego na zachowanie człowieka zostały przedstawione w opracowaniu *E. Gomez* (red.), Assessing the impact of machine intelligence on human behaviour: an interdisciplinary endeavour, Barcelona 2018. Coraz szersze i nietransparentne wykorzystanie rozwiązań opartych o szeroko rozumianą sztuczną inteligencję powodują, że ludzie stają się nieświadomie „niewolnikami" algorytmów, które w połączeniu z dość swobodnym podejście do naszych danych osobowych, może generować realne ryzyko w zakresie uzależnienia się od podmiotów, które mogą dokonywać manipulacji nami samymi w bardzo zróżnicowanym zakresie, w tym w obszarze zakupowym, a także społecznym. Z tego względu rozważne podejście do kwestii etyki sztucznej inteligencji jest bardzo istotne, choć nie można zapominać o innym ważnym aspekcie – rozwoju innowacji i możliwościach, jakie daje wykorzystanie modeli SI, np. w wykrywaniu i leczeniu wielu schorzeń. Na tym ostatnim polu również pojawiają się rozwiązania regulacyjne, np. WTO, Ethics and Governance of Artificial Intelligence for Health. WHO Guidance 2021, https://apps.who.int/iris/rest/bitstreams/1352854/retrieve (dostęp z 26.10.2021 r.).

[8] *S.M. Liao* (red.), Ethics of Artificial…, s. 25–26.

[9] *J. Basl*, *J. Bowen*, AI as a Moral Right-Holder, [w:] *M.D. Dubber*, *F. Pasquale*, *S. Das*, The Oxford Handbook of Ethics of AI, Oxford 2021, s. 289 i n.

[10] Koncepcję tę dobrze opisuje *B. Goertzel*, Artificial General Intelligence: Concept, State of the Art, and Future Prospects, Journal of Artificial General Intelligence 2014, Nr 5(1), s. 1–46. Artykuł, jak też zawarte w nim rozważania, pomimo że napisane w 2014 r., zachowują swoją aktualność. Samo pojęcie tzw. AGI można sprowadzić do próby stworzenia sztucznej inteligencji najbardziej zbliżonej do inteligencji ludzkiej, czyli nakierowanej na szerszą analizę otoczenia i wyciąganie wniosków z tego otoczenia, a nie realizowanie pojedynczych zadań wyznaczanych przez człowieka. Konsekwencją stworzenia AGI może być konieczność nadania pewnej podmiotowości takiemu systemowi i powiązanych z tym praw, ale również nauczenie go odpowiednich postaw moralnych. Stworzenie AGI, przynajmniej w najbliższej przyszłości, jawi się jednak jako mało prawdopodobne, ale z ostrożności naukowej nie można tego wykluczyć.

woduje to, że rozważania w kontekście rozwoju sztucznej inteligencji na obecnym poziomie technologicznym niekoniecznie dotykają sedna problemu związanego m.in. z prawami podstawowymi i zagrożeniami dla człowieka. Na tę kwestię zwraca m.in. *S. Chesterman*, którego zdaniem powinniśmy skupić się na tych problemach i wyzwaniach, szczególnie o charakterze prawnym, które mają znaczenie dla człowieka już dzisiaj[11].

Akcent w dyskusji na temat etyki sztucznej inteligencji powinien w znacznej mierze ogniskować się wokół roli człowieka[12] w stosowaniu modeli SI, a także zapewnieniu odpowiednich rozwiązań organizacyjnych[13] oraz technicznych w tym obszarze. Jednocześnie nie oznacza to, że dywagacje na temat praktycznej możliwości implementacji norm etycznych w algorytmach czy modelach nie są potrzebne[14] lub całkowicie przedwczesne. Trzeba jednak zwrócić uwagę, że nie to powinno stanowić główną oś tych dyskusji w tym zakresie. Nie można oczywiście wykluczyć, że w najbliższej przyszłości sztuczna inteligencja stanie się bardziej „rozumna" od człowieka[15], tym bardziej że już dziś znacznie przewyższa człowieka w niektórych aspektach[16]. Na dziś jednak wydaje się, że dyskusja ta odwraca uwagę od naprawdę istotnych problemów związanych z coraz większym wykorzystaniem modeli sztucznej inteligencji do manipulacji ludzkimi zachowaniami. Te problemy są zaś w dużej mierze zależne od człowieka, a jednym z głównych zagrożeń nie jest wcale szkoda fizyczna (choć nie jest ona wykluczona), ale raczej psychiczna i ekonomiczna powiązana z zagadnieniem stronniczości algorytmicznej oraz dyskryminacji[17].

Niniejsze opracowanie nie ma na celu wykazania błędu w prowadzonych badaniach nad autonomicznością sztucznej inteligencji i potrzebą tworzenia etycznej SI. Uwaga zostanie skoncentrowana na rozwiązaniach, których opracowanie jest istotne z punktu widzenia obecnego stanu rozwoju modeli czy systemów szeroko rozumianej sztucznej inteligencji, a nie przyszłych (i niepewnych) rozwiązań. Z tego względu rozważania na temat tzw. AGI zostaną ograniczone do niezbędnego minimum, natomiast akcent zostanie w znacznej mierze położony na praktyczne aspekty etyki w kontekście „maszyn" czy bardziej obsługujących je ludzi. Jednocześnie jest to jedynie wstęp do dyskusji nad sposobem realizacji założeń etycznej sztucznej inteligencji i w założeniu niniejszy artykuł stanowi pierwszą część otwierającą serię poświęconą temu zagadnieniu.

Wydaje się, że choć rozwiązania o charakterze technicznym[18] są potrzebne, to jednak nie jest to kierunek w chwili obecnej do końca realny do zrealizowania (a być może nawet optymalny, jeżeli uwzględnić, jakie ograniczenia mogą wiązać się ze stosowaniem takich rozwiązań[19]), co jest konsekwencją:

– braku pełnej wiedzy co do funkcjonowania ludzkiego mózgu[20] (choć to nie wydaje się głównym problemem) i identyfikacji wzorców moralnych czy postępowania[21];
– niepewności co do źródła norm etycznych[22];
– braku jasnego i ujednoliconego katalogu tych norm oraz sposobu ich wartościowania;
– ograniczonego rozwoju technologicznego w zakresie AGI, jak również w zakresie możliwości wypracowania i zastosowania konkretnych narzędzi dla bardziej realnych zastosowań sztucznej inteligencji, jak uczenie maszynowe czy głębokie[23];

[11] *S. Chesterman*, We, the Robots? Regulating Artificial Intelligence and the Limits of the Law, Cambridge 2021, s. 4–5.

[12] *B. Lepri*, *N. Oliver*, *A. Pentland*, Ethical machines: The human-centric use of artificial intelligence, iScience24, 102249, March 19, 2021, https://reader.elsevier.com/reader/sd/pii/S2589004221002170?token=A2093D639FE713344042036C63DC87C82020E534F86DC52AB8FC9B19E23C3F4D256966FDF293DC393186129F0E57462F&originRegion=eu-west-1&originCreation=20211029212417 (dostęp z 30.10.2021 r.).

[13] *M. Maleszak*, *P. Zaskórski*, Systems and Models of Artificial Intelligence in the Management of Modern Organisations, Information Systems in Management 2015, Vol. 4 (4), s. 267 i n.

[14] *V.C. Müller*, Ethics of Artificial Intelligence and Robotics", *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), *E.N. Zalta* (red.), https://plato.stanford.edu/archives/sum2021/entries/ethics-ai/> (dostęp z 30.10.2021 r.).

[15] *F. Amigoni*, *V. Schiaffonati*, Machine Ethics and Human Ethics: A Critical View, https://www.aaai.org/Papers/Symposia/Fall/2005/FS-05-06/FS05-06-016.pdf (dostęp z 30.10.2021 r.).

[16] *A. Stohr*, *J. O'Rourke*, Through the Cognitive Functions Lens – A Socio-Technical Analysis of Predictive Maintenance, 16th International Conference on Wirtschaftsinformatik, March 2021, Essen, Germany, Conference Paper, March 2021.

[17] *E. Ntoutsi*, *P. Fafalios*, *U. Gadiraju* (et. al.), Bias in data-driven artificial intelligence systems – An introductory survey, Wiley Interdisciplinary Reviews: Data mining and knowledge discovery (10)6, May 2020.

[18] Warto tutaj zwrócić uwagę na opracowanie IEEE zob. *R. Chatila* et al., IEEE Global Initiative Aims to Advance Ethical Design of AI and Autonomous Systems, IEEE SPECTRUM, Mar. 2017, https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e.pdf?utm_medium=undefined&utm_source=undefined&utm_campaign=undefined&utm_content=undefined&utm_term=undefined (dostęp z 30.10.2021 r.).

[19] *T. Hagendorff*, The Ethics of AI Ethics: An Evaluation of Guidelines, Minds and Machines 2020, Nr 30, s. 108 i n.

[20] Jest to szczególnie istotne w kontekście dyskusji nad możliwością odwzorowania etycznych wzorców w modelach sztucznej inteligencji. *W. Wallach*, *C. Allen*, *I. Smit*, Machine Morality: Bottom-up and Top-down Approaches for Modeling Human Moral Faculties, https://www.aaai.org/Papers/Symposia/Fall/2005/FS-05-06/FS05-06-015.pdf (dostęp z 5.11.2021 r.).

[21] Problem ten dotyczy tzw. *value alignment*, który *M. Risse* tłumaczy jako konieczność ustalenia, jakie normy i wartości są dla nas rzeczywiście ważne (a także jak je stopniować, jeżeli w ogóle). *M. Risse*, Human Rights and Artificial Intelligence An Urgently Needed Agenda, May 2018, Carr Center for Human Rights Policy, https://carrcenter.hks.harvard.edu/files/cchr/files/humanrightsai_designed.pdf (dostęp z 5.11.2021 r.).

[22] *J. Taylor*, *E. Yudkowsky*, *P. LaVictoire*, *A. Critch*, Alignment for Advanced Machine Learning Systems, [w:] *S. M. Liao* (red.), Ethics of Artificial…, s. 346.

[23] O trudnościach związanych z przełożeniem norm etycznych (czy też prawnych) na modele sztucznej inteligencji pisze m.in. *I. Gabriel*, Artifcial Intelligence, Values, and Alignment, Minds and Machines 2020, Nr 30, s. 413–415. Z kolei *D. Szostek* rekomenduje konkretne rozwiązania związane z wykorzystaniem algorytmizacji do regulowania sztucznej inteligencji. *D. Szostek*, Is the Traditional Method of Regulation (the Legislative Act) Sufficient to Regulate Artificial Intelligence, or Should It Also Be Regulated by an Algorithmic Code?, Białostockie Studia Prawnicze 2021, Nr 3, vol. 26.

– problematyki tzw. wyjaśnialności (*explainability*)[24] oraz ograniczeń tzw. inżynierii odwrotnej (*reverse engineering*).

Jednocześnie trzeba zwrócić uwagę, że nie oznacza to zanegowania potrzeby wypracowania takich rozwiązań, które niewątpliwie w niektórych przypadkach (samochody samojezdne czy platformy społecznościowe) są pożądane. Wydaje się jednak, że obecnie dyskurs naukowy zmierza bardziej w kierunku próby stworzenia wzorca agenta moralnego[25] (dla) sztucznej inteligencji, nie zaś próby realnego rozstrzygnięcia wyzwań związanych z postępującą algorytmizacją ludzkiego życia. To natomiast powoduje, że ucieka nam *clue* problematyki stosowania modeli sztucznej inteligencji, czyli ułatwiania życia człowiekowi i poprawa dobrostanu społecznego przy poszanowaniu praw podstawowych[26].

Jednocześnie na plan pierwszy w kontekście szeroko rozumianej etycznej sztucznej inteligencji wysuwa się nie tyle problematyka wyboru „metody" implementacji norm etycznych, ile samego określenia, względem „kogo" lub „czego" stosujemy określone wymogi. Z tego względu niniejszy artykuł jest poświęcony głównie zagadnieniu pojęcia (systemów) sztucznej inteligencji, które stanowić powinno punkt wyjścia dla dyskusji nad tzw. moralnością maszyn.

Artykuł został podzielony na cztery obszary. Pierwszy stanowi próbę wprowadzenia w problematykę tzw. silnej sztucznej inteligencji oraz wyzwań etycznych. W drugim podjęta zostanie próba odpowiedzi na pytanie o potrzebę definiowania sztucznej inteligencji jako takiej. W kolejnym rozdziale zaproponowane zostanie podejście oparte na definicji systemów sztucznej inteligencji, które może rozwiązać wiele z problemów natury prawnej, ale i etycznej. Na końcu zawarte są natomiast konkluzje oraz postulaty *de lege ferenda*, a także przyczynek do dalszej dyskusji.

Czy ludzkość potrzebuje definicji sztucznej inteligencji?

Jednym z najszerzej dyskutowanych zagadnień w kontekście szeroko rozumianej sztucznej inteligencji jest jej definicja[27]. Pojęcie to i próba znalezienia „optymalnego" rozwiązania są przedmiotem debat[28] na poziomie zarówno naukowym[29], jak i politycznym[30], co u przeciętnego odbiorcy może generować znaczny niepokój i zagubienie. Jest to podyktowane często chęcią zapewnienia pewności prawnej[31] i zminimalizowania ryzyka niepewności po stronie adresatów określonych norm. W przypadku sztucznej inteligencji próba jednoznacznego jej zdefiniowania jest to – paradoksalnie – działanie „sztuczne". Panuje obecnie zgoda to co do tego, że nie wiemy, w którym kierunku zmierzają rozwiązania oparte o szeroko rozumianą sztuczną inteligencję[32], sami nie jesteśmy bowiem w stanie przewidzieć naszych możliwości, a także możliwości metod, które stosujemy, opracowując SI. Warto przy tym zauważyć, że jednym z wyzwań, które będzie miało kluczowe znaczenie w kontekście koegzystencji człowieka i rozwiązań opartych na AI, jest tzw. *augmented intelligence*[33], czyli wzmacniania zdolności poznawczych i kognitywistycznych człowieka.

[24] *A. Holzinger*, *G. Langs*, *H. Denk* (et. al.), Causability and explainability of artificial intelligence in medicine, WIREs Data Mining and Knowledge Discovery (9), 2019, https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/widm.1312 (dostęp z 5.11.2021 r.).

[25] *F. Fossa*, Artificial moral agents: moral mentors or sensible tools?, June 2018, Ethics and Information Technology, 20(1), s. 2. Wydaje się jednak, że bez możliwości przełożenia czy to norm prawnych, czy etycznych na format (przejrzysty) do odczytu maszynowego (*machine readable*) próby stworzenia narzędzia, a nawet fundamentów dla etycznej sztucznej inteligencji nie przyniosą oczekiwanego rezultatu. Jest to pochodną przekonania, że jeżeli coś nie jest „zero-jedynkowe" (abstrahując od tzw. logiki rozmytej – *fuzzy logic*), to nie jest możliwe do skutecznej „egzekucji" przez maszynę, która „myśli" właśnie kategoriami binarnymi. Każda decyzja podjęta przez SI, która będzie wymykała się takiemu jednoznacznemu wzorcowi, nie będzie odpowiadała założeniom etycznej SI, chyba że jako społeczeństwo uznamy, że również maszyny mogą podejmować decyzje oparte na „intuicji" czy innych „miękkich" wartościach. Zagadnienie to zostanie szerzej opisane w dalszej części opracowania.

[26] *A. Kayid*, The role of Artificial Intelligence in future technology, 15.3.2020 r., https://www.researchgate.net/publication/342106972_The_role_of_Artificial_Intelligence_in_future_technology (dostęp z 5.11.2021 r.).

[27] *L. Lai*, *M. Świerczyński*, Prawo sztucznej inteligencji, Warszawa 2020, s. 9.

[28] *D.S. Grewal*, A Critical Conceptual Analysis of Definitions of Artificial Intelligence as Applicable to Computer Engineering, IOSR Journal of Computer Engineering (IOSR-JCE) 2014, Vol. 16, Issue 2, s. 3. *D.S. Grewal* wskazuje, że ostatnie 40 lat to okres gwałtownych zmian w zakresie definicji sztucznej inteligencji, która ewoluuje wraz z rozwiązaniami technologicznymi leżącymi u jej podstaw. *D.S. Grewal* rekomenduje definiowanie SI jako system maszynowej symulacji bazującej na pobieraniu wiedzy oraz informacji, a następnie przetwarzaniu otoczenia i przenoszeniu rezultatu tego przetwarzania na bardziej zrozumiałe działania „inteligentne".

[29] Coraz częściej problematyka ta łączona jest ze wspomnianą przez autora potrzebą utworzenia pojęcia etycznej sztucznej inteligencji, co tylko zwiększa niepewność prawną, ale i wyzwania społeczne związane z tym zagadnieniem – zob. *J. Fjeld*, *A. Nele*, *H. Hilligoss*, *A. Nagy*, *M. Srikumar*, Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI, Berkman Klein Center for Internet & Society 2020, s. 11–13.

[30] *R. Braun*, Artificial Intelligence: Socio-Political Challenges of Delegating Human Decision-Making to Machines, 2019, IHS Working Paper 6.

[31] Pytanie, czy próba twardego regulowania sztucznej inteligencji jest podejściem właściwym. Na całym świecie widoczna jest tendencja do łączenia podejścia opartego na twardych i miękkich regulacjach i szerszym stosowaniu neutralności technologicznej połączonej z podejściem opartym na ryzyku (*risk-based approach*), które z jednej strony przerzuca ryzyko prawne i regulacyjne na podmioty stosujące określone rozwiązania, ale z drugiej daje znacznie większą swobodę i może przyczyniać się do rozwoju innowacji. *R.H. Weber*, Overcoming the Hard Law/Soft Law Dichotomy in Time of (Financial) Crises, Journal of Governance and Regulation 2012, Vol. 1, Issue 1, s. 13–14. Także *R. Hagemann*, *J. Huddlestone Skees*, *A. Thrierer*, Soft Law for Hard Problems: The Governance of Emerging Technologies in an Uncertain Future, Colorado Technology Law Revie 2018, Vol. 17.1, s. 58–59.

[32] *D.A. Bray*, The Future of Artificial Intelligence, https://www.businesofgovernment.org/sites/default/files/Chapter%20Fourteen%20The%20Future%20of%20AI.pdf (dostęp z 8.11.2021 r.).

[33] *I. Kushchu*, *T. Demirel* (red.), Artificial Intelligence Media and Information Literacy, Human Rights and Freedom of Expression, UNESCO 2020, s. 4, https://iite.unesco.org/wp-content/uploads/2021/03/AI_MIL_HRs_FoE_2020.pdf (dostęp z 8.11.2021 r.).

W tym kontekście warto zauważyć, że obecnie rzeczywiście nie ma podstaw do twierdzenia, że GAI jest w ogóle możliwe do wypracowania[34], natomiast przykładów łączenia aspektów technologicznych i ludzkich przybywa. Generować mogą one – i z pewnością będą – wiele wyzwań o charakterze prawnym, etycznym[35] czy filozoficznym, a ich bliskość człowiekowi powoduje, że będą też stanowić dla niego zagrożenie.

Pojawia się więc zasadnicze pytanie – czy powinniśmy definiować na potrzeby prawne sztuczną inteligencję[36]? A jeżeli tak, to w którym kierunku powinna iść ta definicja i na ile powinna pozostawać ona otwarta, a na ile szczegółowa[37], a także rozważyć opcję definiowania w zależności od zastosowania (konkretnej regulacji)[38]. Wydaje się jednak, że powinniśmy się raczej skupić na odpowiedzi na następujące pytanie – czy chcemy regulować „sztuczną inteligencję", czy raczej stosowanie rozwiązań opartych o różne metody i techniki (jak uczenie maszynowe i głębokie), które stanowią część szeroko pojętej dziedziny sztucznej inteligencji?

Co to oznacza w praktyce? Jeżeli przyjrzymy się najnowszym opracowaniom międzynarodowych organizacji oraz instytucji, w tym unijnych, to okaże się, że w większości przypadków odchodzi się od próby definiowania samej sztucznej inteligencji[39] na rzecz tzw. systemów sztucznej inteligencji. Spójrzmy na konkretne przykłady, przy czym nie chodzi tutaj o wskazywanie błędów, które można zidentyfikować w samych definicjach, a jedynie o przedstawienie pewnego (pozytywnego) trendu w tym zakresie.

Systemy sztucznej inteligencji jako odpowiedź

Projekt rozporządzenia Parlamentu Europejskiego i Rady ustanawiającego zharmonizowane przepisy dotyczące sztucznej inteligencji (dalej: AIA)[40] wprowadza definicję „systemu sztucznej inteligencji", przez który rozumie się oprogramowanie opracowane przy użyciu co najmniej jednej spośród technik i podejść wymienionych w załączniku I (do tego projektu – chodzi tutaj m.in. o uczenie maszynowe, metody statystyczne czy przetwarzanie języka naturalnego), które może – dla danego zestawu celów określonych przez człowieka – generować wyniki, takie jak treści, przewidywania, zalecenia lub decyzje wpływające na środowiska, z którymi wchodzi w interakcję. Również OECD definiuje w ramach swoich Zasad[41] „system sztucznej inteligencji" rozumiany jako system oparty na maszynie, który jest w stanie wpływać na środowisko poprzez wytwarzanie danych wyjściowych (przewidywania, zalecenia lub decyzje) dla danego zestawu celów. Wykorzystuje on dane i dane wejściowe pochodzące od człowieka i/lub maszyny w celu: (i) postrzegania rzeczywistych i/lub wirtualnych środowisk; (ii) przetwarzania tego postrzegania w modele poprzez analizę w sposób zautomatyzowany (np. za pomocą uczenia maszynowego) lub ręcznie oraz (iii) wykorzystania wnioskowania z modelu do sformułowania opcji dla możliwych wyników. Systemy SI są zaprojektowane do działania z różnym poziomem autonomii. Podobnie propozycja rekomendacji UNESCO[42] w sprawie etycznej sztucznej inteligencji nie tylko proponuje oparcie się o pojęcie (dość rozbudowane) systemów sztucznej inteligencji, ale też w sposób wyraźny wskazuje na trudności związane z tą definicją, w tym konieczność jej zmiany wraz z rozwojem technologii.

Pewien wyłom można odnaleźć w przypadku rekomendacji Światowej Organizacji Zdrowia[43], gdzie odniesienia można znaleźć zarówno do systemów SI, jak i sztucznej inteligencji, jednak wydaje się, że panuje zgodność, że konieczne jest przede wszystkim uregulowanie kwestii systemów sztucznej

[34] *R.V. Yampolskiy*, *J. Fox*, Artificial General Intelligence and the Human Mental Model, [w:] *A. Eden*, *J. Søraker*, *J.H. Moor*, *E. Steinhart* (red.), Singularity Hypotheses: A Scientific and Philosophical Assessment, Berlin 2012, s. 11–13. Pomimo że artykuł powstał prawie dekadę temu, zmiany w tym zakresie nie nastąpiły w na tyle dużym natężeniu, aby uzasadniały przyjęcie tezy o możliwości zasymulowaniu działania ludzkiego mózgu przez komputer.

[35] *H. Hassani*, *E.S. Silva*, *S. Unger*, *M. TajMazinani*, *S. Mac Feely*, Artificial Intelligence (AI) or Intelligence Augmentation (IA): What Is the Future?, AI 2020, s. 151, https://www.mdpi.com/2673-2688/1/2/8/pdf (dostęp z 8.11.2021 r.).

[36] Problematyce tej bardzo dokładnie przyjrzała się Komisja Europejska, która w opracowaniu z 2020 r. wskazała na trudności, jakie wiążą się z próbą stworzenia jednolitej definicji SI, wskazując m.in. na komponenty, które powinna ona zawierać, tj. wybór kierunku definicji (w tym techniczna versus nietechniczna), taksonomia oraz słowa kluczowe istotne dla poszczególnych poddziedzin SI. Komisja Europejska, AI Watch Defining Artificial Intelligence Towards an operational definition and taxonomy of artificial intelligence 2020, s. 7 i 86.

[37] *P. Wang* wskazuje, że choć mamy wiele definicji sztucznej inteligencji, to żadna z nich nie jest „perfekcyjna". Nie jest to jednak zdaniem tego autora zasadniczy problem i należy oczekiwać, że w najbliższej przyszłości się to nie zmieni – zob. *P. Wang*, On Defining Artificial Intelligence, Journal of Artificial General Intelligence 10(2), 2019, s. 29.

[38] Na wielość znaczenia pojęcia sztucznej inteligencji wskazuje chociażby *P. Wang*, What do you mean by „AI", Frontiers in Artificial Intelligence and Applications, Nr 171(1), s. 365.

[39] Rekomendacje WHO w zakresie etycznej sztucznej inteligencji dla obszaru zdrowia nieco się „wyłamują" z tego ogólnego podejścia, wskazując na to, jak definiuje się „sztuczną inteligencję", choć jednocześnie referują do propozycji definicji wytworzonej przez OECD. WHO, Ethics and Governance of Artificial Intelligence for Health, 2021, https://apps.who.int/iris/rest/bitstreams/1352854/retrieve (dostęp z 9.11.2021 r.).

[40] Projekt rozporządzenia Parlamentu Europejskiego i Rady ustanawiającego zharmonizowane przepisy dotyczące sztucznej inteligencji, COM(2021) 206 final 2021/0106 (COD).

[41] Zob. https://oecd.ai/en/ai-principles (dostęp z 9.11.2021 r.). AI system is a machine-based system that is capable of influencing the environment by producing an output (predictions, recommendations or decisions) for a given set of objectives. It uses machine and/or human-based data and inputs to (i) perceive real and/or virtual environments; (ii) abstract these perceptions into models through analysis in an automated manner (e.g., with machine learning), or manually; and (iii) use model inference to formulate options for outcomes. AI systems are designed to operate with varying levels of autonomy.

[42] UNESCO, Draft text of the Recommendation on the Ethics of the Artificial Intelligence SHS/IGM-AIETHICS/2021/JUN/3 Rev.225, June 2021.

[43] WHO, Ethics and Governance of Artificial Intelligence for Health, 2021, https://apps.who.int/iris/rest/bitstreams/1352854/retrieve (dostęp z 11.11.2021 r.).

inteligencji, a więc systemów (czy oprogramowania), które wykorzystuje różne metody i techniki stanowiące szeroko rozumianą sztuczną inteligencję oddziałujących na człowieka i jego środowisko. Należy przy tym podkreślić, że dobrym rozwiązaniem byłoby przy tym pozostawienie poza sferą definicji włączanie tych metod i podejść, jako niebędących (jedynym) niezbędnym elementem składowym samej sztucznej inteligencji. Przywoływana już definicja systemów sztucznej inteligencji połączona z zestawem metod i podejść[44], określona w AIA budzi wiele kontrowersji ze względu na jej niezwykle szerokie ramy, które powodują, że „zwykłe" oprogramowanie niemające charakteru samouczącego się również może być kwalifikowane jako sztuczna inteligencja[45]. Prowadzi to do sytuacji, w której nie tylko wprowadza się pewne zamieszanie terminologiczne, ale także poddaje surowym i specyficznym wymogom prawnym rozwiązania będące daleko od sfery uczenia maszynowego czy głębokiego. Warto tutaj przywołać opinię *M. Świerczyńskiego* i *Z. Więckowskiego*, którzy podkreślają, że „(…) łatwiejsze do definiowania jest pojęcie systemu sztucznej inteligencji niż samej sztucznej inteligencji"[46].

Takie podejście ma zasadniczą zaletę, że nie próbuje definiować czegoś (kogoś?), co nie jest w tej chwili do jednoznacznego zdefiniowania, pozostawiając tym samym pewien margines interpretacji, ale jednocześnie wprowadza pewne ograniczenia (zasadne) prawne dla rozwiązań, które zbliżają się do szerokiego rozumienia SI, co ma jedną zasadniczą – choć zupełnie nienaukową – wartość, czyli promowanie świadomości społecznej w zakresie tego rozwijającego się obszaru technologii, mającego istotny wpływ na człowieka. Zagadnienie to jest jednak poza zakresem niniejszego opracowania, choć wydaje się niezwykle istotne w kontekście szeroko rozumianej edukacji. Jednocześnie *M. Fischer* oraz *S. Parab*[47] wskazują, iż pojęcie SI jest tak szerokie (i pojemne), że trudno jest znaleźć jedną (dobrą) definicję, tym bardziej że ewolucja rozwiązań w tym obszarze jest znacząca, choć nadal daleka od wspomnianej już GAI. Tym samym powinniśmy – choć nie zapominając o zastrzeżeniach poczynionych przez *S. Russela*[48] – nie tyle skupić się na definiowaniu sztucznej inteligencji i próbie jej „uczłowieczania", ile położyć akcent na takim tworzeniu rozwiązań o charakterze prawnym, regulacyjnym i społecznym (oraz infrastrukturalnym), które zapewnią, że SI będzie tworzone w sposób etyczny i w taki sposób będzie też działać, ale nie w znaczeniu autonomicznym, o czym w dalszej części opracowania.

W tym miejscu należy poczynić jeszcze jedną uwagę, choć zagadnienie wykracza znacząco poza zakres artykułu. Na poziomie UE toczy się obecnie dyskusja nad definicją oprogramowania[49] i jej wpływem na SI, w szczególności czy sztuczna inteligencja powinna stanowić część „oprogramowania", czy też ze względu na swój charakter powinna podlegać wyłączeniu z tej definicji. Jest to zagadnienie o tyle istotne, że brak jednoznacznego zakwalifikowania przykładowo systemów SI do ww. definicji spowodować może istotne wątpliwości interpretacyjne, a także nadmierne lub zbyt słabe wymagania. Niewątpliwie jednak, ze względu na swoją specyfikę, sztuczna inteligencja powinna stanowić pewien podzbiór oprogramowania o wyraźnie wyodrębnionych ramach prawnych (w tym wymogach w zakresie danych), co jest pochodną jej istotnego wpływu na człowieka, przede wszystkim w kontekście praw podstawowych. W tym miejscu nie zostanie jednak rozstrzygnięte, jakie podejście wydaje się najwłaściwsze, choć patrząc na propozycję definicji systemów sztucznej inteligencji, a więc właśnie „wyodrębnienie" SI jako podzbioru oprogramowania, wydaje się na dzisiaj najwłaściwsze.

Konkludując tę część opracowania, należy stwierdzić, że obecnie nie jest pożądane, ani niezbędne, definiowanie samej sztucznej inteligencji, ale szeroko rozumianych ram jej zastosowania, czyli wspomnianych już systemów sztucznej inteligencji. Oczywiście należy mieć na uwadze, że nie eliminuje to nam stanu pewnej niepewności[50], w dalszym ciągu bowiem nie wiemy, czym jest sztuczna inteligencja. Wspomniana już definicja systemu sztucznej inteligencji ma jedną zasadniczą wadę – próbuje stworzyć kompletną listę podejść czy technik z zakresu szeroko rozumianej SI – co powoduje, że z jednej strony mamy tam rzeczywiście rozwiązania samouczące się (bliskie przynajmniej potocznemu rozumieniu SI), jak też

[44] Na istotne różnice w podejściach i technikach z szeroko rozumianego obszaru sztucznej inteligencji wskazują chociażby *N. Kuhl*, *M. Goutier*, *R. Hirt*, *G. Satzger*, Machine Learning in Artificial Intelligence: Towards a Common Understanding, Proceedings of the 52nd Hawaii International Conference on System Sciences 2019, https://core.ac.uk/download/pdf/211327717.pdf (dostęp z 17.11.2021 r.). Szczególnie istotne w tym kontekście jest wyraźne rozróżnienie tych rozwiązań, które mają charakter „uczących się", od tych, które działają wyłącznie według wcześniej ustalonych „standardów", wyszukując przykładowo korelacje. Dystynkcja ta jest o tyle istotna, że spora część praktyków SI wskazuje, iż o sztucznej inteligencji (jeżeli w ogóle) można mówić w przypadku systemów samouczących się.

[45] Na tę problematykę wskazują m.in. *P. Langley* oraz *J.E. Laird*, Artificial Intelligence and Intelligent Systems, https://www.researchgate.net/publication/250150471_Artificial_Intelligence_and_Intelligent_Systems (dostęp z 14.11.2021 r.).

[46] *M. Świerczyński*, *Z. Więckowski*, Sztuczna inteligencja w prawie międzynarodowym. Rekomendacje wybranych rozwiązań, Warszawa 2021, s. 39.

[47] *M. Fischer*, *S. Parab*, Regulating AI. What Everyone Needs to Know about Artificial Intelligence and the Law, Self-Replicating AI Press 2020, s. 21.

[48] *S. Russel*, Artificial Intelligence. A Binary Approach, [w:] *S.M. Liao* (red.), Ethics of Artificial…, s. 327 i n.

[49] *Ch. Wendechorst*, Safety and Liability Related Aspects of Software, Luksemburg 2021, https://ec.europa.eu/newsroom/dae/redirection/document/77327 (dostęp z 17.11.2021 r.).

[50] Również Parlament Europejski oraz eksperci zaangażowani w proces konsultacji nad projektem rozporządzenia w sprawie SI zwracają uwagę na szerokie ramy zaproponowanej definicji systemów sztucznej inteligencji. Podsumowanie z listopada 2021 r. – zob. https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf (dostęp z 18.11.2021 r.).

„klasyczne" metody statystyczne[51], które jakkolwiek mogą wpływać (negatywnie) na człowieka, nie powinny być utożsamiane z SI.

Rozwiązaniem w tym zakresie byłoby pozostawienie poza sferą definicyjną zakresu możliwych podejść, a bardziej ogólne określenie „charakteru" tych metod i podejść, np. poprzez następującą zmianę (przy czym taka definicja mogłaby być wykorzystana nie tylko na potrzeby wspomnianego projektu rozporządzenia): „system sztucznej inteligencji" oznacza oprogramowanie wykorzystujące dostępne i znane techniki samodzielnego uczenia się [przez to oprogramowanie], które posiada funkcjonalność pozwalającą – dla danego zestawu celów określonych przez człowieka – generować (ustalone przez człowieka) wyniki, takie jak treści, przewidywania, zalecenia lub decyzje wpływające na środowiska zewnętrzne, na które takie wyniki mogą oddziaływać zarówno w sposób pozytywny, jak i negatywny, w tym poprzez wyrządzenie szkody.

Powyższa definicja zawiera oczywiście pewne niejasne pojęcia, które mogą być różnie interpretowane, co zawsze będzie budziło (i budzi) wątpliwości, głównie wśród prawników preferujących „sztywne" definicje i ograniczenie roli tzw. *soft law*. W szczególności należy tutaj zwrócić uwagę, że przyjęcie szerokiej definicji systemów sztucznej inteligencji może mieć ten skutek, że niektóre rozwiązania „dalekie" od SI (czy uczenia maszynowego lub głębokiego[52]) będą w taki sposób definiowane. Zwrócić jednak także należy uwagę na fakt, że samo wprowadzenie takiej definicji do aktu prawnego nie musi jeszcze wiązać się z nałożeniem konkretnych wymagań w tym zakresie, czego dowodem jest chociażby wspomniany już projekt rozporządzenia. Oczywiście wymaga to pewnej kategoryzacji systemów (np. ze względu na poziom generowanego ryzyka) oraz konsekwencji w tym zakresie. Warto jednocześnie pamiętać o istotnej kwestii, czyli roli, jaką odgrywa miękkie prawo w kształtowaniu otoczenia prawno-regulacyjnego, w szczególności w świecie dynamicznie zmieniających się technologii. Rozwiązania prawne dotykające tematyki nowych (nowoczesnych) technologii muszą wykracza poza nieco „konserwatywne" podejście prezentowane w większości aktów prawnych. Oczywiście wiąże się to z pewną dozą niepewności, która może być jednak równoważona rozsądnymi i efektywnymi rozwiązaniami z obszaru miękkiego prawa.

Nie jest to definicja doskonała, jednak przy obecnym stanie wiedzy technologicznej[53], wielości aktów „około" sztucznej inteligencji (w tym braku jednolitego pojęcia oprogramowania oraz rozwoju różnych technik) prawdopodobnie każda próba uporządkowania tego pojęcia będzie obarczona ryzykiem „niedoprecyzowania" czy zbytniej ogólności. Pojawia się więc pytanie, czy powinniśmy w dalszym ciągu próbować stworzyć perfekcyjną definicję systemów sztucznej inteligencji, czy też skupić się bardziej na klarownym i zrównoważonym obudowaniu tych systemów stosownymi wymogami, które będą stanowiły swoisty konsensus pomiędzy prawnikami a osobami „technologicznymi".

Wydaje się, że przyjęcie tego drugiego podejścia ma znacznie większą wartość, obecnie bowiem największym wyzwaniem jest ochrona praw jednostki, która w związku z rozwojem różnych produktów i usług wykorzystujących algorytmy i modele sztucznej inteligencji połączonych z dość dużą swobodą, z jaką ludzie udostępniają dane o nich samych, może być istotnie nadwątlona. Przeciwdziałać temu powinny odpowiednie rozwiązania prawne, a ściślej ich poprawna implementacja oraz efektywny nadzór nad spełnianiem poszczególnych wymogów. Klasyfikacja określonych systemów jako będących lub niebędących systemami sztucznej inteligencji zawsze będzie sprawiała pewne trudności, tak jak zawsze arbitraż regulacyjny i próby obchodzenia przepisów. W takim przypadku, jeśli zapewnione zostaną odpowiednie ramy instytucjonalne, zawsze dopuszczalna będzie ingerencja odpowiedniego organu nadzorczego. Brak szczególnych i specyficznych wymagań dla takich systemów (w szczególności o wysokim stopniu ryzyka dla człowieka i jego otoczenia) powoduje, że jednostki pozostają niejako bezbronne wobec działań podmiotów wdrażających takie rozwiązania.

Konkluzje i postulaty *de lege ferenda*

W niniejszym opracowaniu dokonano jedynie zarysowania problematyki związanej ze sztuczną inteligencją i jej regulacją. Dyskusja nad definicją SI stanowi jedynie jeden z elementów globalnej dyskusji nad jej przyszłością, która obejmuje takie zagadnienia jak wyjaśnialność (*explainability*)[54] i audytowalność modeli, stronniczość algorytmiczna i zagadnienie dyskryminacji, jak również tworzenie etycznej (godnej zaufania) sztucznej inteligencji. W szczególności ten ostatni wątek może rozpalać wyobraźnię, dotyka bowiem wielu sfer, które można określić jako co najmniej sporne, tj. wybór właściwego katalogu norm (czy takowy w ogóle istnieje?), przełożenia tych norm na konkretne wymagania – najpierw prawne, a następnie techniczne czy wreszcie problematykę rozstrzygania odpowiedzialności za „działania" sztucznej inteligencji.

[51] *J. Lu*, Statistical methods with applications to machine learning and artificial intelligence, 2012, praca doktorska: https://smartech.gatech.edu/bitstream/handle/1853/44730/lu_yibiao_201208_phd.pdf (dostęp z 26.11.2021 r.).

[52] Nawet ta najbardziej zaawansowana metoda czy technika szeroko rozumianej sztucznej inteligencji ma określone ograniczenia i wady, przykładowo w obszarze identyfikacji obrazów, na co wskazują chociażby *B. Zochuri*, *M. Moghaddam*, Deep Learning Limiations and Flaws, Modern Approaches on Material Science 2020.

[53] Warto tutaj zwrócić uwagę, jak stosunkowo niewiele zmieniło się w koncepcji sztucznej inteligencji od chwili jej „utworzenia". *R. Chrisley*, The Development of the Concept of Artificial Intelligence Historical Overviews and Milestones 2000, Artificial Intelligence: Critical Concepts.

[54] *G. Bar*, Przejrzystość, w tym wyjaśnialność, jak wymóg prawnych dla systemów Sztucznej Inteligencji, MoP 2020, Nr 20, s. 75 i n.

Dzisiaj powinniśmy skupić się na budowaniu silnych ram prawnych oraz instytucjonalnych dla tych rozwiązań, które funkcjonują obecnie, a nie zostaną (o ile w ogóle) wynalezione za kilkadziesiąt lub kilkaset lat. Jakkolwiek można wyobrazić sobie, że taka silna sztuczna inteligencja nas zaskoczy, to jednak przyjąć należy, że będziemy w stanie „zareagować" odpowiednio szybko na zagrożenia z nią związane. Skupić należy się więc na eliminowaniu zagrożeń, które są „tu i teraz", tym bardziej że coraz częściej w przestrzeni publicznej pojawiają się informacje o nadużyciach ze strony twórców platform wykorzystujących SI. Jednym z kluczowych elementów, który otwierać będzie drogę do budowania etycznej SI, jest niewątpliwie jednoznaczne zdefiniowanie systemów sztucznej inteligencji.

Istotne jest przy tym wyraźne podkreślenie, że próba nadania „etycznego" charakteru samej sztucznej inteligencji, czyli *de facto* algorytmom i modelom, może stanowić ślepy zaułek dla rozwoju zarówno samej technologii, jak i rozwiązań prawnych. Obecnie kluczowe jest bowiem określenie obowiązków „aktorów" związanych z tworzeniem tego typu rozwiązań oraz ich wykorzystywaniem w przestrzeni publicznej, a także jednoznaczne ustalenie zakresu (i kierunku) odpowiedzialności za działanie poszczególnych rodzajów (kategorii) systemów sztucznej inteligencji. Jeżeli bowiem akcent zostanie położony głównie na obszar związany z implementacją norm etycznych przez same „maszyny", nie rozwiąże to problemu kompensacji ewentualnych szkód, która stanowić powinna następstwo nieetycznego, ale nie nieodpowiedniego w sensie technicznym, działania systemów SI.

Z tego względu, zarówno na poziomie UE, jak i krajowym, rekomendowanymi rozwiązaniami będą:

1) uregulowanie pojęcia systemów sztucznej inteligencji oraz powiązanie tej definicji (lub wykluczenie z zakresu) z pojęciem oprogramowania, a także wprowadzenie kategorii ryzyka, jakie takie systemy mogą generować;
2) jednoznaczne uregulowanie zakresu obowiązków oraz odpowiedzialności poszczególnych podmiotów znajdujących się w łańcuchu funkcjonowania SI; jednym z kierunków, który wydaje się godnym rozważenia przynajmniej w odniesieniu do systemów SI „dotykających" człowieka, jest podejście oparte na ryzyku; jednocześnie należy tutaj rozsądnie podejść do kwestii udowodnienia ewentualnej szkody spowodowanej działaniem systemu SI;
3) ustalenie pewnego katalogu norm etycznych (kodeksów), na których systemy SI powinny się opierać; punktem wyjścia mogą być tutaj prawa podstawowe stanowiące uniwersalny katalog norm, natomiast należy również rozważyć normy o charakterze „sektorowym", które mogą mieć specyficzne znaczenie oraz wartość np. w branżach regulowanych, jak sektor finansowy czy motoryzacyjny;
4) rozważenie wprowadzenia wymogów udziału specjalisty z zakresu etyki (sztucznej inteligencji) w procesie projektowania, tworzenia i wdrażania systemów sztucznej inteligencji (wysokiego ryzyka), tzw. *ethics by default and design*;
5) ugruntowanie charakteru rozwiązań z zakresu etyki o charakterze *self-governance* i konsekwencji niezastosowania się do wewnętrznych norm (lub sektorowych), np. na wzór Badania i Oceny Nadzorczej (BION) stosowanego m.in. w sektorze bankowym – wymaga to jednak równoczesnego ustalenia organu odpowiedzialnego za nadzór nad SI.

Jednocześnie niniejszy artykuł należy traktować jako wstęp do dyskusji, która toczy się nad etyką sztucznej inteligencji. Docelowe wypracowanie rozwiązania w zakresie tworzenia, ale i nadzorowania etycznej SI, której poprawne działanie ma istotne znaczenie z perspektywy całego społeczeństwa, jest bowiem bardzo ważne.

Słowa kluczowe: sztuczna inteligencja, etyka, uczenie głębokie, agent moralny, zasady.

Moral responsibility of HAL 9000, i.e., the ethics of artificial intelligence in practice. Do we need a definition of artificial intelligence?

Recently, we have seen the development of various methods in the area of the so-called artificial intelligence (AI), including the consideration of ethical aspects in the context of new technologies. This causes, at least for today, not necessarily effective discussions and considerations, the practical value of which, in the context of the development of AI on the current technological level, seems to be small. At the same time, what comes to the fore in the context of broadly defined ethical artificial intelligence is not so much the issue of choosing a "method" of implementing ethical standards, as the very definition of "whom" or "what" in regard to applying given requirements. Therefore, this article is mainly devoted to the concept of artificial intelligence (systems), which should be the starting point for the discussion of the so-called morality of machines.

Key words: artificial intelligence, ethics, deep learning, moral agent, principles.