

Tomasz Matuszek – Uniwersytet Śląski

## Sztuczna inteligencja a człowieczeństwo

### Streszczenie

Poniższy artykuł zawiera pogłębioną analizę współczesnej koncepcji człowieka i jego zderzenia z problematyką urządzeń sztucznej inteligencji. Pogłębionej refleksji krytycznej zostało poddane nazewnictwo stosowane w określaniu sztucznej inteligencji, określono problem humanizowania urządzeń i dehumanizowania człowieka. Koniecznym okazał się zabieg odwołania do charakterystyki osoby i jej bycia.

Czy urządzenia sztucznej inteligencji mogą być szansą na zwrot ku przywróceniu człowiekowi jego człowieczeństwa? Jak można zdefiniować odpowiedzialność maszyny jako urządzenia stworzonego przez człowieka? Obserwując narastające napięcie między zwolennikami *maszynizacji* i *demaszynizacji* świata można wyciągnąć wniosek, że staje się nie nad przepaścią apokalipsy ze strony techniki, lecz nad przepaścią świadomości *bycia człowiekiem*. Niniejszy artykuł postara się przekrojowo podjąć tematykę człowieczeństwa w obliczu urządzeń sztucznej inteligencji oraz dowieść, że wspomniane napięcie może być szansą na redefinicję *człowieczeństwa*.

W obliczu postępującego rozwoju urządzeń sztucznej inteligencji – SI (ang. *artificial intelligence, AI*) – niezależna grupa ekspertów wysokiego szczebla ds. sztucznej inteligencji powołana przez Komisję Europejską (w skrócie HLEG AI) opublikowała *Wytyczne w zakresie etyki dotyczącej godnej zaufania sztucznej inteligencji*<sup>1</sup>. Czy można zatem zaufać maszynie? Jeśli tak, to czy maszyna może zawieść zaufanie lub czy można traktować ją po partnersku? Jest to jednak raczej niebezpieczny zabieg językowy, który może zaburzyć pewne rozumienie *rzeczywistości* jako przestrzeni, którą rozumie i kreuje przede wszystkim – jeśli nie tylko – człowiek. „Ostatecznie SI nie powinna być przedstawiana jako godna zaufania, ponieważ podważa to wartość zaufania międzyludzkiego, antropomorfizuje SI (afektywną stronę zaufania) i odrzuca odpowiedzialność tych, którzy wynaleźli i korzystają z SI (normatywna

---

<sup>1</sup> Komisja Europejska, Dyrekcja Generalna ds. Sieci Komunikacyjnych, Treści i Technologii, *Wytyczne w zakresie etyki dotyczącej godnej zaufania sztucznej inteligencji*, Bruksela 2019.

strona zaufania)<sup>2</sup>. Tym samym nie należy postawić znaku równości między sztuczną inteligencją a inteligencją człowieka.

Istnieje jednak ryzyko uproszczenia tematu, dlatego nie można pominąć założeń, które towarzyszą rozróżnieniom między *weak AI* (słabą SI, która stosuje wgrane algorytmy) a *strong AI* (mocną SI, która potrafi tworzyć nowe – nieznane wcześniej algorytmy, które nie pochodzą z ewolucji wgranych)<sup>3</sup>, ponieważ nawet jeśli weźmie się pod uwagę *wersję strong*, to należy podkreślić, że nie jest ona tożsama z *człowieczeństwem*. Istotna wydaje się tutaj nomenklatura, gdyż urządzenie pozostaje urządzeniem („mechanizm lub zespół mechanizmów, służący do wykonania określonych czynności”<sup>4</sup>). Nie można zapominać, że to człowiek nadaje cechy technologii, nie odwrotnie. Czy istnieje zatem jakaś wyjątkowość w człowieku, jakiś warunek konieczny do tego, by uznać jego *bycie człowiekiem*? Odhumanizowanie urządzeń sztucznej inteligencji staje się coraz częściej przedmiotem dyskusji nie tylko w ramach debat naukowych, ale także podczas codziennych rozmów. Prowadzone są badania nad zaufaniem społecznym wobec urządzeń SI<sup>5</sup>. Humanizacja coraz częściej zaczyna dotyczyć samych ludzi, a nie być jedynie bezpodstawną antropomorfizacją zwierząt, roślin czy właśnie urządzeń. Paradoksalnie zatem, to właśnie urządzenia SI mogą stać się szansą dla redefinicji i nowego odkrycia człowieczeństwa.

Temat ten, dlaczego człowiek jest człowiekiem, towarzyszy rozważaniom filozoficznym od wieków. Jednak dotychczasowe odpowiedzi przestają wystarczać, bowiem zarówno człowiek, jak i urządzenia SI potrafią tworzyć, działać wspólnie, rozróżniać pojęcia i się uczyć. W ostatnim stuleciu personalisci założyli, że człowiek jest osobą odpowiedzialną za siebie i świat w spotkaniu z sobą i drugim. Pojawia się więc także pytanie o to, czy urządzeniu SI należy przypisać odpowiedzialność, chociażby prawną<sup>6</sup>. Czy możliwe jest, żeby maszyna – gdy popełni błąd – stanęła w prawdzie? Za przykład może posłużyć sytuacja, z jaką

<sup>2</sup> Tłum. Własne, oryg.: „Ultimately, AI should not be viewed as trustworthy because it undermines the value of interpersonal trust, anthropomorphises AI (the affective account of trust), and diverts responsibility from those developing and using AI (the normative account of trust)”. Por. M. Ryan, *In AI We Trust: Ethics, Artificial Intelligence, and Reliability*, „Sci Eng Ethics” 26 (2020), s. 2749-2767.

<sup>3</sup> S. Russel, P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed., New Jersey 2010, s. 1020.

<sup>4</sup> *Słownik Języka Polskiego*, PWN; [sjp.pwn.pl/slowniki/urz%C4%85dzenie.html](http://sjp.pwn.pl/slowniki/urz%C4%85dzenie.html) [dostęp: 20.06.2023r.].

<sup>5</sup> N. Gillespie, S. Lockey, C. Curtis, J. Pool, A. Akbari, *Trust in Artificial Intelligence: A Global Study*, Brisbane 2023.

<sup>6</sup> Zob. P. Stylec-Szromek, *Sztuczna inteligencja – prawo, odpowiedzialność, etyka*, Zeszyty Naukowe Politechniki Śląskiej. Seria: organizacja i zarządzanie (z. 123), Wydawnictwo Politechniki Śląskiej 2018; M. Czochoła, D. Bar, *Śmierć pacjenta wywołana zastosowaniem sztucznej inteligencji w technologiach medycznych – analiza prawnokarna*, *Studia Prawnicze. Rozprawy i Materiały* 2019, nr 2 (25); R. Rejmaniak, *Autonomiczność systemów sztucznej inteligencji jako wyzwanie dla prawa karnego*, *Roczniki Nauk Prawnych* t. XXXI, nr 3 (2021), s. 95-113; D. Kaczan, *Odpowiedzialność odszkodowawcza za funkcjonowanie sztucznej inteligencji w medycynie*, „*Studia de Cultura*” 14(2) 2022.

spotkał się mój kolega, który niedawno zapytał o coś ChatGPT i następnego dnia, już po otrzymaniu odpowiedzi, okazało się, że Chat *przeprószył* wspomnianego kolegę, gdyż *kłamał* i podał błędne informacje. Co to jednak znaczy, – że Chat kłamał? Miał *świadomość* i *celowo* wprowadził w błąd? Tego raczej się nie dowiemy, jednak analizując aktualny poziom rozwoju ChatuGPT, można raczej uznać, że jest to efekt braku danych, a nie autorefleksji. Poza tym, czy mamy pewność, że nie jest to wynik wprowadzonego programu, żeby Chat *uwierzytelnić* zachowaniami *ludzkimi*? Nie można także zapomnieć, że zachowania ludzkie nie decydują o *byciu człowiekiem*. Przecież osoba będąca w stanie wegetatywnym nie może kłamać i przeprzeć, a jednak niepodważalnie jest człowiekiem.

Stajemy zatem wobec gigantycznego dylematu nad istotą *człowieczeństwa*. Viktor Frankl w książce *Homo patiens* uznaje, że to *możność* do stawania się definiuje człowieczeństwo<sup>7</sup>, jednak nie można zapomnieć, że „nie chodzi tylko o to, aby być człowiekiem, ale o to, aby się stać samym sobą”<sup>8</sup>. Tym samym stawia się istotną różnicę między maszynową aktualizacją, która po wykryciu błędów *naprawia je* i nie dopuszcza do powtórzenia. Człowiek się bowiem nie aktualizuje, człowiek nie przestanie popełniać błędów, ale człowiek wkracza w głąb siebie do tego stopnia, że odkrywa swoją *niepowtarzalność*. Jeśli zatem każdy jest niepowtarzalny, to jednocześnie staje się w *najgłębszym* tego słowa znaczeniu *niewyraźalny*. Tego zatem nie można odmówić nawet osobie w stanie wegetatywnym. Owszem, można kogoś pozbawić odpowiedzialności prawnej i moralnej<sup>9</sup>, ale nie można pozbawić tej *niezwykłości*. Musimy *wyłączyć* tutaj wszelką działalność człowieka *ad extra*, musimy wykluczyć jakikolwiek *czyn*. Musimy wykluczyć także *cel*. Aktualna dyskusja bowiem przekracza relacyjność człowieka w wymiarze jego uzewnętrzniania lub przeżywania w sposób wyraźny. Daleko idąca humanizacja maszyn, w tym przypadku maszyn SI, stawia wyzwanie, które nie może zatrzymać się na poziomie fizycznym. Sytuacja wymaga, niestety, przekroczenia poziomu *meta-fizycznego*. Nawet bowiem, jeśli dojdzie do działalności *maszyn SI strong*, to właśnie tutaj (sic!) zetknie się rozróżnienie *człowieczeństwa* i *maszynowości*. Nie w sposobie wyrażania, ale w sposobie *bycia* (Niewątpliwie dopiero bycie można wyrazić, ale nie jest to związek konieczny. Równie dobrze można być i się nie wyrażać, i nie być, a coś wyrażać).

Pojawia się tutaj jednak pewien problem, otóż równie dobrze maszyna może być niewyraźalna lub niepowtarzalna. Przecież i ona może mieć *świadomość bycia*. Co więcej,

<sup>7</sup> V.E. Frankl, *Homo patiens*, Warszawa 1984, s. 61-62.

<sup>8</sup> Tamże, s. 64.

<sup>9</sup> *Kodeks postępowania cywilnego*, Art. 544-560, (Dz.U. z 2020 r. poz. 1575 z późn. zm.).

maszyna zawsze będzie sprawna fizycznie i zawsze będzie w najwyższym procesie poznawczym lub w możliwości do zaktualizowania się, w przeciwieństwie do dziecka lub do osoby chorej, słabszej lub potrzebującej. Tutaj znowu powraca przytoczony wyżej problem stawania się<sup>10</sup>, jednak konieczne wydaje się pewne uzupełnienie. Otóż nie można uzależnić człowieczeństwa od możliwości, to samo przecież można powiedzieć o *inteligentnej, uczuciowej, otwartej na zmiany* maszynie.

Zwracając jednak, dla formalności, uwagę na owo *stawanie się* w perspektywie maszyny, należy podkreślić, że wykrywając błąd zacznie [maszyna] go usuwać, zacznie go korygować i zmieniać *na przyszłość*, ponieważ wprowadzone algorytmy lub ich samodzielne modyfikacje będą się uczyły<sup>11</sup>. Zaiste tutaj wydaje się kluczowe rozróżnienie – maszyna poprzez *stawanie się* będzie docierać do coraz mocniejszego zrozumienia siebie i wyrażenia siebie – będzie musiała wykonywać pewne procesy, zadania – pewną naukę. Człowiek natomiast od początku *bytuje* i nie musi się rozwijać, dojrzewać lub wzrastać, żeby być. To, czy człowiek się wyraża, czy nie, jest tylko pewną naturalną konsekwencją – w dodatku niekonieczną. Jeżeli przytoczy się maszynę, która jest zaprogramowana tak, że *odgrywa* stan wegetatywny, to po pierwsze będzie kawałkiem metalu lub zbiorem bajtów, a po drugie ewidentnie będzie to wynik jakiegoś zamierzonego działania, czyli pewien sposób wyrażenia siebie – siebie wegetatywnego.

Czy w pełni sprawną fizycznie, osobę chorą psychicznie, uznamy za nie-człowieka? Przecież nie mamy z nią kontaktu, a jednak potrafimy stworzyć pewne systemy, które ułatwiają z nią kontakt (Nazizm uznawał pewne grupy za *podludzi*, a to jednak dalej *jacyś* ludzie godni lub nie godni życia – ten problem jednak wydaje się rozwiązany w Norymberdze). Tak samo osoba w stanie wegetatywnym jest nieustannie uznawana za człowieka – nawet szeroko pojęty problem eutanazji tak naprawdę jest w znaczącej mierze sporem etycznym, rozbija się on bowiem o zabójstwo i jego możliwość, nie o to, czy zabijamy jeszcze człowieka, czy już nie, bo z zasady dotyczy to żyjącego (dramatem jest to, że człowiek nie chce żyć!). Stawać się może i człowiek, i maszyna, ale tylko człowiek może być, bo tylko człowiek *musi* być. Niepodważalne wydaje się tutaj zatem *bycie* od samego początku do samego końca, od poczęcia do śmierci. Maszyna też *jest*, ale tylko dlatego, że się jakkolwiek wyraża. Dopiero teraz człowiek może się stawać – gdy już *jest*. Maszyna może być, bo się staje, ale nigdy nie będzie, bo proces stawania jest nieskończony.

<sup>10</sup> V. E. Frankl, *Homo...*, s. 61-64.

<sup>11</sup> S. Russel, P. Norvig, *Artificial...*, s. 693-694.

Spoglądanie na *bycie* otwiera pewną tezo-hipotezę co do istoty człowieczeństwa, że człowiek w zderzeniu z urządzeniami SI dostrzeże różnicę w przeżywaniu i odkrywaniu swojej egzystencji w aspektach i wymiarach, które nierzadko są niezależne od niego samego, a jednak to one rozróżniają go od urządzeń SI. Dalszym pytaniem niewątpliwie winno być to, *skąd* byt człowieka *jest*, co również może wskazywać na pewną znaczną różnicę w podejmowanym temacie – jesteśmy bowiem w stanie podać konkretne daty powstania pierwszych urządzeń sztucznej inteligencji – ona nie została odkryta – nie jest nowym gatunkiem, ale została stworzona w konkretnym momencie czasoprzestrzennym, o którym wiadomo. Pojawienie się *urządzeń SI* jest szansą dla człowieka, aby zdefiniował siebie, aby na nowo odkrył swoją niepowtarzalność. Natomiast to, co z nią zrobić, doskonale podjęli personalisci – Tischner, Wojtyła, Frankl, Stein, czy wcześniej Ingarden. Bowiem gdy już *jesteśmy*, to chociaż nie musimy, to *powinniśmy się stawać*.

### **Bibliografia**

CZOCHRA M., BAR D., *Śmierć pacjenta wywołana zastosowaniem sztucznej inteligencji w technologiach medycznych – analiza prawnokarna*, Studia Prawnicze. Rozprawy i Materiały 2019, nr 2 (25).

FRANKL V.E., *Homo patiens*, Warszawa 1984.

GILLESPIE N., LOCKEY S., CURTIS C., POOL J., AKBARI A., *Trust in Artificial Intelligence: A Global Study*, Brisbane 2023.

KACZAN D., *Odpowiedzialność odszkodowawcza za funkcjonowanie sztucznej inteligencji w medycynie*, „Studia de Cultura” 14(2) 2022.

*Kodeks postępowania cywilnego*, Art. 544-560, (Dz.U. z 2020 r. poz. 1575 z późn. zm.).

KOMISJA EUROPEJSKA, DYREKCJA GENERALNA DS. SIECI KOMUNIKACYJNYCH, TREŚCI I TECHNOLOGII, *Wytyczne w zakresie etyki dotyczącej godnej zaufania sztucznej inteligencji*, Bruksela 2019.

REJMANIAK R., *Autonomiczność systemów sztucznej inteligencji jako wyzwanie dla prawa karnego*, Roczniki Nauk Prawnych t. XXXI, nr 3 (2021), s. 95-113. RUSSEL S., NORVIG P., *Artificial Intelligence: A Modern Approach*, 3rd ed., New Jersey 2010.

RYAN M., *In AI We Trust: Ethics, Artificial Intelligence, and Reliability*, „Sci Eng Ethics” 26 (2020).

*Słownik Języka Polskiego*, PWN; [sjp.pwn.pl/slowniki/urz%C4%85dzenie.html](http://sjp.pwn.pl/slowniki/urz%C4%85dzenie.html) [dostęp: 20.06.2023r.].

STYLEC-SZROMEK P., *Sztuczna inteligencja – prawo, odpowiedzialność, etyka*, Zeszyty Naukowe Politechniki Śląskiej. Seria: organizacja i zarządzanie (z. 123), Gliwice 2018.