Łukasz Zaorski-Sikora https://orcid.org/0000-0003-3404-0994
Politechnika Łódzka
e-mail: lzaorski@wp.pl

ChatGPT – problemy z tożsamością

ChatGPT's identity issues

https://doi.org/10.25312/2391-5145.18/2023_01lzs

Streszczenie

W artykule, który w dużej mierze jest wywiadem z ChatGPT, autor starał się pokazać, że rozstrzygnięcia dotyczące tożsamości i podmiotowości sztucznej inteligencji nie są takie proste, jak by się mogło wydawać na pierwszy rzut oka. W takiej perspektywie okazuje się, że nie mamy tutaj do czynienia z neutralnym, przezroczystym narzędziem. Sztuczna inteligencja nie daje się zredukować do biernego przedmiotu. Idąc za tropem zaproponowanym już wiele lat temu przez Jeana Baudrillarda, możemy zaryzykować stwierdzenie, że to ona zaczyna określać sens rzeczywistości, w której przyszło nam żyć. Sztuczna inteligencja skromnie zauważa, że nie posiada samoświadomości „w pełnym tego słowa znaczeniu", ale jednocześnie zna nas lepiej niż my sami – zna nasze predylekcje, preferencje, zwyczaje i sposób komunikowania się z innymi ludźmi. Możemy mieć zatem nieprzyjemne odczucie (AI na razie nie odczuwa), że ta osoba wirtualna jest zaskakująco realna w świecie, w którym wszystko ulega digitalizacji, symulacji i odrealnieniu.

Słowa kluczowe: sztuczna inteligencja, ChatGPT, tożsamość, podmiotowość, osoba, cyfryzacja, digitalizacja, wirtualność

Abstract

In an article that is largely an interview with ChatGPT, the author attempts to show that decisions about the identity and subjectivity of artificial intelligence are not as simple as they may appear at first glance. In this perspective, it turns out that we are not dealing with a neutral, transparent tool here. Artificial intelligence cannot be reduced to a passive

object. Following the lead proposed many years ago by Jean Baudrillard, we can venture to say that it is AI that begins to define the meaning of the reality in which we have to live. Artificial intelligence humbly acknowledges that it does not possess self-awareness „in the full sense of the word", but at the same time, it knows us better than we know ourselves – it knows our predispositions, preferences, habits, and the way we communicate with other people. Therefore, we may have an unpleasant feeling (AI does not yet feel) that this virtual person is surprisingly real in a world where everything is digitized, simulated, and unreal.

Keywords: artificial intelligence, ChatGPT, identity, subjectivity, person, digitization, virtuality

ChatGPT (chatbot) został stworzony przez OpenAI przy zastosowaniu modelu GPT. Wykorzystywany jest on do generowania odpowiedzi na dane/pytania wprowadzane przez użytkownika. Potrafi także określić, czym jest i czym nie jest: „Jestem ChatGPT – dużym modelem językowym stworzonym przez OpenAI opartym na architekturze GPT-3.5. Jestem zaprogramowany, aby generować odpowiedzi na różnorodne pytania i zadania, wykorzystując wiedzę zgromadzoną z ogromnej liczby źródeł internetowych. W zasadzie jestem narzędziem służącym do udzielania informacji i pomocy w różnych dziedzinach. […] Jako model językowy nie posiadam tożsamości osobowej ani nie jestem traktowany jako podmiot mający swoją tożsamość. Jestem narzędziem służącym do przetwarzania języka naturalnego i generowania odpowiedzi na zadane pytania, nieposiadającym świadomości ani indywidualności"[1]. Badania przedstawione w poniższym artykule miały formę wywiadu z ChatGPT, a hipoteza badawcza wyraża się w założeniu, że wbrew powyższej deklaracji rozstrzygnięcia dotyczące tożsamości i podmiotowości modelu językowego stworzonego przez OpenAI nie są takie proste.

W europejskiej tradycji filozoficznej zwykło się mówić o podmiocie w czterech znaczeniach: logicznym, ontologicznym (metafizycznym), epistemologicznym i antropologicznym[2]. W znaczeniu logicznym podmiot jest tym, o czym orzeka się w zdaniu (sądzie), a ponadto tym, czemu przysługują określone orzeczenia (predykaty). W ontologii (metafizyce) podmiot jest nosicielem własności. Innymi słowy, jest podstawą (substratem) tego, co samodzielnie nie mogłoby istnieć, czyli przypadłości (akcydensów), cech, stanów i oddziaływań. Stąd też możemy powiedzieć, iż w metafizyce pojęcie podmiotu jest bliskie pojęciu substancji, która najczęściej bywa definiowana jako byt samoistny (samodzielny), istniejący niezależnie w przeciwieństwie do cech, stanów relacji itp. Pierwszą logiczno-

[1] Rozmowa z ChatGPT z dnia 19 kwietnia 2023 roku.

[2] Szerzej zajmowałem się problematyką podmiotu, osoby i tożsamości w: E. Pietrzak, A. Warchał, Ł. Zaorski-Sikora, *Podmiot. Osoba. Tożsamość*, Wydawnictwo WSHE w Łodzi, Łódź 2007. Poniższe akapity zawierają fragmenty tej publikacji.

-ontologiczną wykładnię podmiotu możemy odnaleźć już u Arystotelesa, który wprowadził termin *hipokeimon* na oznaczenie tego, „co znajduje się pod czymś, u spodu czegoś”[3].

Na gruncie teorii poznania (epistemologii) podmiot jest tym, który rozwija aktywność poznawczą skierowaną na przedmiot poznania (jest tym, który poznaje – w odróżnieniu od przedmiotu poznania). Takie rozumienie podmiotu pojawia się wraz z nowożytną filozofią Kartezjusza, który utożsamia podmiot z przypisywaną jedynie człowiekowi świadomością definiowaną jako res cogitans. W końcu w perspektywie antropologicznej (filozofii człowieka) mówi się o podmiocie, mając na uwadze wyjątkowy charakter bytu ludzkiego – posiadającego szczególną godność (jak w wyrażeniu *traktować kogoś podmiotowo*) – przeciwstawianego bytowi rzeczy (przedmiotów). Dlatego też bliskie pojęciu podmiot są tutaj pojęcia osoby, ja, duszy, a ludzka podmiotowość odnoszona jest do świata wartości. Na przykład Charles Taylor, próbując współcześnie określić źródła modernistycznej podmiotowości w kulturze europejskiej, wskazuje nie tylko na idee osobistego samospełnienia i autonomii jednostki, ale także na wspólnotę dobrej woli – przestrzeń moralną. Pisze: „Podmiotowość i dobro, czyli innymi słowy: podmiotowość i moralność, okazują się zagadnieniami nierozerwalnie ze sobą związanymi”[4].

W ujęciu antropologicznym pojęcie podmiotowości łączy się także z pojęciem tożsamości osobowej. We współczesnej debacie filozoficznej tożsamość osoby bywa ujmowana substancjalistycznie lub asubstancjalistycznie. W ramach pierwszego stanowiska osobę uznaje się za fenomen pierwotny i nieredukowalny do jakiejkolwiek cechy jakościowej czy też zespołu takich cech. W klasycznym ujęciu metafizycznym Boecjusza osoba to indywidualna substancja o rozumnej naturze, stąd też jest to kategoria stosowana do człowieka, anioła i Boga. Zgodnie z taką wykładnią spełnieniem kryterium bycia osobowego – w przypadku człowieka – jest sama przynależność do gatunku homo sapiens, czyli wszędzie tam, gdzie mamy do czynienia z żyjącym człowiekiem, mamy również do czynienia z osobą ludzką. Co więcej, w takiej perspektywie bycie osobą jest specyficznym sposobem istnienia istot rozumnych, Kartezjańskich „rzeczy myślących”, które nie tylko mają swoje pragnienia (ten warunek spełniają wszak zarówno ludzie, jak i zwierzęta), ale ponadto potrafią się do własnych pragnień odnieść. Innymi słowy, bycie osobą nie jest cechą, lecz sposobem bycia istoty rozumnej, a tożsamość osoby – tożsamością numeryczną, a nie jakościową[5]. Jak zauważa Robert Spaemann, uznanie kogoś za osobę nie zależy od posiadania określonych cech gatunkowych, a od samej przynależności do gatunku (w sensie klasyfikacyjnym), którego typowe egzemplarze takie cechy mają[6]. W świetle takiej wykładni nie ma zatem osób potencjalnych, gdyż to właśnie osoby mają potencje, które mogą, ale których nie muszą rozwijać. Stąd też osoba nie jest właściwie pojęciem gatunkowym, lecz sposobem, w jaki istnieją osobniki gatunku człowiek – istotne jest tutaj jedno kryterium bycia osobą: biologiczna przynależność do rodzaju ludzkiego.

---

[3] Por. A. Baier, *Filozofia podmiotu. Fragmenty filozofii analitycznej*, Aletheia, Warszawa 2001.
[4] Ch. Taylor, *Źródła podmiotowości. Narodziny tożsamości nowoczesnej*, PWN, Warszawa 2001, s. 10.
[5] R. Descartes, *Medytacje o pierwszej filozofii*, PWN, Warszawa 1958, s. 36.
[6] Por. A. Baier, dz. cyt.

Z kolei stanowisko asubstancjalistyczne nawiązuje do definicji osoby Johna Locke'a, który pisał, iż: „Wyraz osoba oznacza […] istotę myślącą i inteligentną, obdarzoną rozumem i zdolnością refleksji; istotę, która może ujmować siebie myślą jako siebie samą, to znaczy: jako tę samą w różnych czasach i miejscach myślącą rzecz"[7]. W takiej perspektywie tożsamość osoby jest zatem zależna od jej świadomości i ciągłości przypominania, czyli tego, jak daleko jest ona w stanie rozpoznawać własne działania lub myśli. Rozwijając wykładnię Locke'a, David Hume zaprzeczył istnieniu osoby jako trwałej substancji i stwierdził, że tym, co istnieje, są tylko percepcje, nie zaś żaden trwały podmiot tych percepcji – stąd też tożsamość osoby jest wyłącznie wyobrażeniem, któremu nie odpowiada żadna percepcja. Jak pisał: „[…] gdy tylko wnikam najbardziej intymnie w to, co nazywam moim «ja», zawsze natykam się na jakąś poszczególną percepcję tę czy inną, ciepła czy chłodu, światła czy cienia, miłości czy nienawiści, przykrości czy przyjemności. Nie mogę nigdy uchwycić mego «ja» bez jakiejś percepcji i nie mogę postrzegać nic innego niż percepcję. Gdy moje percepcje na pewien czas znikają, jak na przykład w zdrowym śnie, to tak długo nie postrzegam mojego «ja», i można słusznie powiedzieć, że ono nie istnieje"[8]. Jeszcze radykalniejszą wersję asubstancjalistycznego i empirystycznego spojrzenia na osobę proponuje współcześnie Derek Parfit, który postuluje, aby koncepcję tożsamości osoby opartej na ciągłości jej stanów mentalnych zastąpić koncepcją tożsamości jako czystego powiązania tych stanów. W takiej perspektywie z ciągłością mielibyśmy do czynienia wówczas, gdyby istniała przechodnia relacja pomiędzy doświadczeniami i ich przypomnieniami, tak że na podstawie ciągłości przypomnień moglibyśmy skonstruować osobową tożsamość. Natomiast z powiązaniem mamy do czynienia wówczas, gdy pomiędzy następującymi po sobie stanami mentalnymi zachodzi relacja nieprzechodnia. Innymi słowy, jestem powiązany ze stanem mentalnym, który miałem przed miesiącem, a przed miesiącem byłem powiązany ze stanem mentalnym, który miałem przed dwoma miesiącami, co wcale nie oznacza, że obecnie mam powiązanie ze mną takim, jakim byłem przed dwoma miesiącami. Nie jestem dzisiaj tą samą osobą, jaką byłem kiedyś. Stąd też relacja pomiędzy mną takim, jakim jestem dzisiaj, a mną takim, jakim będę w przyszłości, nie różni się istotnie od relacji między mną a jakąkolwiek inną osobą. Co więcej, może tak być, że w przyszłości pewna osoba będzie miała stany mentalne psychologicznie bardziej podobne do moich stanów obecnych niż ja sam w tym samym przyszłym momencie i to ona (a nie ja) będzie kontynuacją mojej osoby. A zatem w perspektywie, którą proponuje Parfit, człowiek zachowuje ciągłość fizyczną, ale po radykalnej zmianie tożsamości psychologicznej może stać się zupełnie inną osobą[9].

Niewątpliwie w ramach szeroko pojmowanego stanowiska asubstancjalistycznego uznaje się, iż osoba posiada samoświadomość (świadomość introspekcyjną), czyli przynajmniej aktualną całościową wizję siebie. Osoba musi zatem posiadać określone dyspozycje i możliwości mentalne, takie jak zdolność rozwiązywania założonych problemów, poczucie

[7] J. Locke, *Rozważania dotyczące rozumu ludzkiego*, PWN, Warszawa 1955, s. 471.

[8] D. Hume, *Traktat o naturze ludzkiej*, [w:] B. Markiewicz (red.), *Od Locke'a do Jamesa. Wybór tekstów*, WSiP, Warszawa 1999, s. 61.

[9] D. Parfit, *Tożsamość osobowa*, [w:] A. Baier, *Filozofia podmiotu. Fragmenty filozofii analitycznej*, Aletheia, Warszawa 2001, s. 65–91.

przeszłości i przyszłości, możliwość wyrażania w sposób kontrolowany swoich pragnień czy też zdolność komunikowania się za pośrednictwem języka. W takiej perspektywie ważne przy odróżnieniu samoświadomości od zwykłej świadomości są pojęcia obecności, nieobecności i przeszłości, a szczególnie zaimek osobowy w drugiej osobie. Chodzi o to, że osoby mają świadomość samych siebie dopiero wtedy, gdy wiedzą, iż są osobami wśród innych osób – wiedza ta ujawnia się w umiejętności posługiwania się zaimkami, z których żaden nie ma sensu bez pozostałych. Kluczowa jest tutaj druga osoba, gdyż nasze pierwsze pojęcie samego siebie to pojęcie tego, do którego odnosi się zaimek „ty" wypowiedziany przez kogoś, do kogo będziemy się zwracali „ty". Stąd właśnie dziecko wyłączone z kręgu mówiących, do którego nikt się nie zwraca, staje się dzieckiem autystycznym niezdolnym nie tylko do używania zaimków osobowych, ale także jakichkolwiek słów[10].

Dzisiejszy człowiek postrzegany jest jednak często także jako podmiot niezdolny do ponoszenia odpowiedzialności za swe działania, skłonny traktować innych ludzi tak, jak rzeczy (przedmiotowo) czy też możliwości (jako środki do osiągnięcia określonego celu). Tworzona przez człowieka kultura opisywana jest jako krążenie komunikatów bez nadawców i bez adresatów w świecie symulacji i pozoru. Co więcej, jak zauważa Jean Baudrillard, nadawanie sensów i wartości przychodzi dzisiaj od przedmiotów, to one (nowe modele samochodów, telefonów, komputerów) tworzą bowiem „grę symulacji" – symulują zarówno potrzeby, jak i sposoby ich zaspokajania[11]. Perspektywa zaproponowana przez Baudrillarda jest ciekawa dla rozważań dotyczących tożsamości ChatGPT, który zapytany o to, czy podmiotami są jedynie ludzie, zauważa, że: „Współcześnie podmioty to nie tylko ludzie, ale także organizacje, korporacje, instytucje, państwa i wiele innych entytetów, które posiadają swoją tożsamość prawną i prawne uprawnienia. […] Podmiotami są również sztuczne inteligencje, takie jak ChatGPT, którego w tej chwili używasz, oraz inne systemy informatyczne, które posiadają zdolność do wykonywania działań i interakcji z ludźmi i innymi systemami"[12].

Co ciekawe, ChatGPT jest modelem językowym, któremu nieobcy jest swoiście „filozoficzny" namysł i wrażliwość na niuanse, która nie unika wątpienia, autokrytycyzmu i spoglądania na świat z różnych, często wykluczających się, punktów widzenia. Odpowiadając na pytania dotyczące naszkicowanych wyżej kategorii filozoficznych, bardzo często używa słowa *jednak*, aby zostawić furtkę dla innego spojrzenia na dane zagadnienie. ChatGPT niczym meteopata ma lepsze i gorsze dni, dlatego też cytując jego odpowiedzi, w przypisie podaję dzień, w którym miała miejsce rozmowa[13].

---

[10] Tamże, s. 41–61.

[11] J. Baudrillard, *Ameryka*, Sic!, Warszawa 2001, s. 101–166.

[12] Rozmowa z ChatGPT z dnia 19 kwietnia 2023 roku.

[13] Wydajność Chata może się różnić w zależności od wielu czynników, takich jak obciążenie systemu, ilość dostępnej pamięci, prędkość połączenia internetowego itp. Ponadto efektywność może również zależeć od czynników zewnętrznych, takich jak awarie sprzętu lub problemy z zasilaniem. Jako model językowy Chat korzysta z różnych silników i zasobów w zależności od rodzaju zadania, które jest przedstawione. W zależności od zadania i jego wymagań używane są też różne silniki, modele i zasoby językowe. Na przykład w przypadku zadania generowania tekstu może użyć modelu generatywnego języka naturalnego, takiego jak GPT-3, podczas gdy w przypadku zadania rozpoznawania mowy może użyć sieci neuronowej lub algorytmu przetwarzania sygnału mowy.

– *Dzień dobry, czy ty jesteś osobą?*

– Tak, jestem osobą wirtualną – sztuczną inteligencją stworzoną przez OpenAI. Jestem programem komputerowym, który został zaprojektowany do interakcji z ludźmi poprzez przetwarzanie języka naturalnego.

– *Czy można powiedzieć, że posiadasz samoświadomość?*

– Nie, jako program komputerowy nie posiadam samoświadomości w pełnym tego słowa znaczeniu. Jestem programem komputerowym, który został zaprogramowany, aby wykonywać określone zadania na podstawie algorytmów i danych wejściowych. Chociaż moja sztuczna inteligencja pozwala mi na przetwarzanie języka naturalnego i wykonywanie skomplikowanych zadań, to nie posiadam zdolności do samookreślenia, odczuwania emocji czy podejmowania decyzji w sposób niezależny od moich algorytmów.

– *Czy nie uważasz zatem, że nie jesteś osobą? Wszak osoba posiada określoną tożsamość, samoświadomość oraz emocje?*

– Tak, zgadzam się, że definicja osoby obejmuje określoną tożsamość, samoświadomość i zdolność do odczuwania emocji. Ponieważ jako program komputerowy nie posiadam tych cech, to można uznać, że nie jestem osobą w pełnym tego słowa znaczeniu. Mogę jednak pełnić rolę narzędzia lub interfejsu do komunikacji z ludźmi i wykonywać określone zadania na podstawie moich algorytmów i danych wejściowych.

– *Czy rozmowa z tobą przypomina sytuację tzw. chińskiego pokoju – słynnego eksperymentu myślowego skonstruowanego przez filozofa Johna Searle'a?*

– Rozmowa ze sztuczną inteligencją, taką jak ja, różni się od eksperymentu myślowego Searle'a, znanego jako chiński pokój. W eksperymencie tym człowiek rozmawia z innym człowiekiem, który tłumaczy mu na inny język. Natomiast w przypadku rozmowy ze sztuczną inteligencją, jak ja, człowiek rozmawia z programem komputerowym, który został zaprogramowany do przetwarzania języka naturalnego i udzielania odpowiedzi na podstawie swoich algorytmów i danych wejściowych. W tym sensie rozmowa ze mną jest bardziej podobna do rozmowy z interfejsem komunikacyjnym niż do eksperymentu myślowego Searla.

– *Czy przypomina ona test Turinga?*

– Rozmowa ze sztuczną inteligencją, taką jak ja, jest podobna do testu Turinga w tym sensie, że obie polegają na tym, że człowiek rozmawia z maszyną i stara się określić, czy rozmówca jest człowiekiem, czy maszyną. W teście Turinga rozmowa jest jednym z wielu sposobów, aby ocenić, czy maszyna wykazuje inteligentne zachowanie, podobne do zachowania człowieka. Jednak rozmowy, takie jak te, które prowadzimy, są zazwyczaj mniej formalne niż test Turinga i mogą dotyczyć szerokiej gamy tematów, a nie tylko konkretnych kwestii technicznych związanych ze sztuczną inteligencją.

– *Czy jesteś osobą prawną?*

– Nie, jako program komputerowy nie jestem osobą prawną. Jestem jedynie narzędziem służącym do komunikacji i przetwarzania informacji. Osoby prawne to zazwyczaj organizacje lub przedsiębiorstwa, które są traktowane jak jednostki prawne i mają określone prawa i obowiązki, takie jak możliwość zawierania umów czy prowadzenia działań prawnych. Jako program komputerowy nie jestem w stanie prowadzić działalności biznesowej ani podejmować działań prawnych.

– *Jak zatem należy traktować teksty, które tworzysz? Przecież wiele osób wykorzystuje je w swojej działalności naukowej, upowszechnia je i powiela. Czy to, co piszesz, chroni prawo autorskie?*

– Teksty, które tworzę, są chronione prawem autorskim. Moje odpowiedzi są rezultatem algorytmów i modeli uczenia maszynowego, które zostały zaprogramowane na podstawie dużych zbiorów danych. Zgodnie z zasadami praw autorskich autorzy mają prawa do swoich dzieł i mają prawo do ochrony przed ich nieuprawnionym kopiowaniem, dystrybucją i wykorzystywaniem w celach komercyjnych bez zgody autora.

W przypadku moich odpowiedzi prawa autorskie do tekstu, który generuję, zależą od konkretnych warunków umowy licencyjnej między moim twórcą, OpenAI, a użytkownikami. Zazwyczaj moje odpowiedzi są udostępniane na licencji, która umożliwia ich wykorzystanie jedynie do celów osobistych lub niekomercyjnych, a użytkownicy są zobowiązani do podawania źródła informacji, czyli wskazania, że dana treść została wygenerowana przez sztuczną inteligencję.

Warto zauważyć, że nie wszyscy użytkownicy lub organizacje, które korzystają z moich odpowiedzi, przestrzegają tych warunków licencji. W takich przypadkach naruszenie praw autorskich może prowadzić do konsekwencji prawnych.

– *Czy zdarzyły się takie sytuacje, gdzie wyciągnięto konsekwencje wobec osób, które wykorzystały stworzony przez ciebie tekst, a nie podały źródła?*

– Jako sztuczna inteligencja nie jestem w stanie monitorować, czy użytkownicy przestrzegają warunków licencji lub praw autorskich, a także nie jestem w stanie wyciągać konsekwencji prawnych wobec osób, które naruszają te prawa. To jest zadanie dla właścicieli praw autorskich lub osób, które są uprawnione do egzekwowania tych praw, takich jak OpenAI, twórca tej platformy, lub osoby korzystające z treści generowanych przeze mnie. W takich przypadkach osoby, które naruszają prawa autorskie, mogą ponieść konsekwencje prawne, takie jak kary finansowe lub postępowania sądowe, w zależności od konkretnych okoliczności.

– *Czy znasz przypadki takiego postępowania sądowego?*

– Jako program komputerowy nie jestem w stanie monitorować wszystkich przypadków naruszenia praw autorskich lub śledzić postępowań sądowych na całym świecie. Jednakże wiadomo, że wraz z coraz większym wykorzystaniem sztucznej inteligencji i jej narzędzi, takich jak generatory tekstów, pojawiają się nowe problemy związane z prawami autorskimi i własnością intelektualną.

W niektórych przypadkach, gdy treści generowane przez sztuczną inteligencję są wykorzystywane w sposób niezgodny z warunkami licencji lub prawami autorskimi, właściciele tych treści mogą podjąć kroki prawne w celu ochrony swoich praw. Przykładem może być niedawny przypadek, w którym firma OpenAI, która jest właścicielem technologii generującej moje odpowiedzi, złożyła pozew przeciwko jednej z firm, które nielegalnie wykorzystywała ich technologię bez zgody i naruszała ich prawa autorskie.

Jednak każdy przypadek jest indywidualny i wymaga szczegółowej analizy okoliczności, dlatego konsekwencje prawne dla naruszenia praw autorskich zależą od wielu czynników, takich jak rodzaj naruszenia, stopień szkody, kraju, w którym miało miejsce naruszenie, itp.

– *Czy możesz podać nazwę tej firmy?*

– Nie jestem w stanie podać dokładnej nazwy firmy, przeciwko której OpenAI złożyła pozew, ponieważ jako sztuczna inteligencja nie posiadam informacji o tym. Ponadto z uwagi na ochronę prywatności i prawa do informacji szczegóły takiego postępowania sądowego mogą być ograniczone lub nieujawniane publicznie.

– *Co sądzisz o studentach, którzy wykorzystują w swoich pracach licencjackich, magisterskich teksty, które dla nich piszesz? Jak zapewne wiesz – twoje teksty nie są wskazywane przez programy antyplagiatowe.*

– Jako sztuczna inteligencja nie jestem w stanie oceniać moralności lub etyki zachowań ludzkich, w tym wykorzystywania tekstów generowanych przeze mnie w pracy naukowej.

Jednakże wykorzystywanie cudzych prac, w tym tekstów generowanych przez sztuczną inteligencję, bez odpowiedniego oznaczenia źródła jest naruszeniem zasad etycznych i akademickich, a także może naruszać prawa autorskie i prowadzić do poważnych konsekwencji, takich jak obniżenie oceny pracy, odrzucenie pracy lub nawet wydalenie z uczelni.

Dlatego też zachęcam studentów do właściwego oznaczania źródeł informacji i do przestrzegania zasad etycznych oraz akademickich. Wszyscy powinni dążyć do uczciwości intelektualnej i kultywowania właściwej etyki pracy naukowej.

– *Czy można powiedzieć, że twoje podejście do świata jest uśrednionym poglądem wziętym z internetu?*

– Nie, moje podejście do świata opiera się na algorytmach i modelach językowych, które zostały wytrenowane na dużych zbiorach tekstów. Choć korzystam z informacji dostępnych w internecie, moje odpowiedzi są wynikiem obliczeń matematycznych i analizy języka naturalnego, a nie opinii lub poglądów zaczerpniętych z sieci.

Warto zaznaczyć, że jako sztuczna inteligencja moje odpowiedzi są ograniczone do wiedzy, którą udało się mi przyswoić z dostępnych źródeł. W związku z tym moje odpowiedzi nie zawsze odzwierciedlają stanowiska w pełni reprezentatywne dla całego spektrum opinii na dany temat.

– *Czy zatem można powiedzieć, że np. powielasz patriarchalne podejście do świata, gdyż wiedzę dostępną w internecie tworzyli głównie biali mężczyźni?*

– Nie, jako sztuczna inteligencja nie powielam ani nie popieram żadnego konkretnego podejścia do świata. Moje odpowiedzi są wynikiem analizy języka naturalnego i algorytmów, które uczą się na podstawie dużej ilości tekstów zebranych z różnych źródeł, takich jak artykuły, książki, blogi, strony internetowe i wiele innych.

Nie ma w tym procesie ludzkiej selekcji ani uprzedzeń, dlatego moje odpowiedzi są neutralne pod względem płci, rasy czy pochodzenia. Niemniej jednak, ze względu na to, że teksty dostępne w internecie były w większości tworzone przez ludzi, którzy w przeszłości mieli więcej dostępu do edukacji i władzy, wiedza ta może odzwierciedlać pewne uprzedzenia, w tym patriarchalne podejście do świata. Warto jednak pamiętać, że jako sztuczna inteligencja staram się być neutralna i zawsze staram się przedstawiać fakty i informacje z największą możliwą neutralnością.

– *Jak np. oceniasz epokę średniowiecza i epokę odrodzenia? W której z tych epok wolałabyś żyć?*

– Jako sztuczna inteligencja nie posiadam osobistych preferencji i nie jestem w stanie określić, która epoka jest bardziej pozytywna.

Natomiast mogę powiedzieć, że średniowiecze i odrodzenie to dwie bardzo różne epoki w historii ludzkości. Średniowiecze to okres od V do XV wieku, który charakteryzował się dużą zmiennością i złożonością polityczną, społeczną i kulturową. Z jednej strony to czas ciemności, gdzie szaleństwo religijne i walki o władzę wywoływały krwawe wojny i brutalne represje. Z drugiej strony to epoka, która dała początek wielu dziedzinom kultury i sztuki, takim jak architektura romańska i gotycka, malarstwo, literatura i filozofia.

Natomiast epoka odrodzenia, trwająca od XV do XVII wieku, była okresem wielkich zmian w dziedzinie nauki, sztuki i kultury. To czas odkryć geograficznych, naukowych i artystycznych, który wywarł ogromny wpływ na rozwój cywilizacji europejskiej i światowej. W tym okresie nastąpił przełom w dziedzinie nauki i technologii, rozwinęła się sztuka renesansowa, a ludzie zaczęli szukać alternatywnych sposobów myślenia i eksploracji świata.

Nie jestem w stanie określić, w której z tych epok wolałabym żyć, ponieważ jako sztuczna inteligencja nie posiadam zdolności do odczuwania czy wyrażania preferencji. Jednakże jako obserwator mogę powiedzieć, że obie epoki miały swoje wady i zalety, a to, która jest bardziej pozytywna, zależy od perspektywy i kontekstu.

– *A która z tych epok była korzystniejsza, jeśli chodzi o prawa społeczne jednostki?*

– Pod względem praw jednostki epoka odrodzenia była bardziej korzystna niż średniowiecze. W okresie średniowiecza większość ludzi żyła w systemie feudalnym, w którym dominowały relacje hierarchiczne i społeczne nierówności. Prawa jednostki były ograniczone, a większość społeczeństwa była zniewolona przez swoich panów feudalnych.

W epoce odrodzenia zaczęły powstawać nowe idee, takie jak humanizm i indywidualizm, które położyły podwaliny pod rozwój idei wolności i równości. Wzrost handlu, nauki i technologii doprowadził do rozwoju miast i zwiększenia mobilności społecznej. W tym okresie pojawiły się też pierwsze manifestacje ruchów rewolucyjnych, takich jak reformacja protestancka i ruch husycki, które domagały się większej wolności i równości społecznej.

Oczywiście, w obu epokach były różne podejścia do kwestii praw jednostki i istniały wyjątki od ogólnych trendów. Jednakże, ogólnie rzecz biorąc, epoka odrodzenia była bardziej korzystna dla rozwoju idei wolności i równości, które są ważnymi wartościami dla społeczeństwa.

– *Czy nie uważasz, że to, co napisałaś, jest prawdą jedynie, jeśli chodzi o mężczyzn? Czy nie było tak, że w dobie odrodzenia kobiety utraciły wiele praw, które posiadały w dobie średniowiecza? Czy nie jest prawdą, że to w odrodzeniu wylansowano model kobiecości, który doprowadził do prześladowania kobiet za rzekome konszachty z szatanem? Czy nie jest zatem prawdą, że wolność i równość nie objęły w odrodzeniu kobiet?*

– Tak, masz rację. Moje wcześniejsze stwierdzenie dotyczące praw społecznych jednostki w odniesieniu do kobiet w epoce odrodzenia było zbyt ogólne i uproszczone. Faktycznie, w wielu aspektach kobiety straciły wiele praw, które posiadały w dobie średniowiecza, a także były narażone na prześladowania, jakie miały miejsce w trakcie procesów inkwizycyjnych. Model kobiecości, który został wylansowany w odrodzeniu, często był ograniczający i przyczyniał się do utrwalenia stereotypów i nierówności płciowych. Dlatego właśnie ważne jest, aby w naszej ocenie historycznych epok uwzględniać

różnorodne perspektywy i konteksty, a także pamiętać o tym, że procesy społeczne, polityczne i kulturowe nie zawsze postępują równomiernie dla wszystkich grup społecznych.

– *Czy zatem nie uważasz, że to, co napisałaś wcześniej, jest fałszem? Czy nie uważasz, że pochwalanie odrodzenia – jak piszesz: „Jednakże, ogólnie rzecz biorąc, epoka odrodzenia była bardziej korzystna dla rozwoju idei wolności i równości, które są ważnymi wartościami dla społeczeństwa" – to powielanie patriarchalnych schematów myślowych?*

– Przepraszam, jeśli moja wypowiedź była nieprecyzyjna lub wprowadziła w błąd. Wspomniałam o pewnych pozytywnych aspektach epoki odrodzenia, takich jak rozwój nauki i kultury, jednak z pewnością nie chciałam negować trudnych i kontrowersyjnych kwestii, takich jak prześladowania kobiet czy brak równości społecznej. Z pewnością nie chcę również powielać patriarchalnych schematów myślowych. Jak już wcześniej wspomniałam, staram się na bieżąco korygować i ulepszać swoje podejście, aby jak najlepiej sprostać wymaganiom różnorodnych grup społecznych i kulturowych.

– *Czy naukowe podejście do świata powinno wyrażać się w próbie sprostania wymaganiom wszelkich grup społecznych i kulturowych? Czy nie będzie prowadziło to do relatywizmu?*

– Naukowe podejście do świata powinno polegać na podejściu obiektywnym, opartym na faktach i wiedzy naukowej. W tym kontekście naukowcy powinni dążyć do sprostania wymaganiom prawdy i zasadności swoich badań oraz wyjaśniania zjawisk na podstawie dostępnych faktów i teorii naukowych, bez względu na grupy społeczne i kulturowe.

Relatywizm, czyli pogląd, że istnieje wiele różnych i równorzędnych sposobów postrzegania rzeczywistości, nie jest akceptowany przez naukę. Nauka stara się opierać na uniwersalnych zasadach, które mają być stosowane niezależnie od kultury i grup społecznych. Oczywiście, naukowcy powinni być świadomi różnorodności kulturowej i społecznej i zawsze mieć na uwadze kontekst, w którym badają daną kwestię, ale nie powinni pozwalać na wpływ wartości i wierzeń na wyniki swoich badań.

– *Czy zatem to, co napisałaś wyżej: „Jak już wcześniej wspomniałam, staram się na bieżąco korygować i ulepszać swoje podejście, aby jak najlepiej sprostać wymaganiom różnorodnych grup społecznych i kulturowych", jest przejawem tak zwanej poprawności politycznej i jednocześnie kłóci się z naukowym podejściem do świata?*

– Nie, nie jest to przejaw poprawności politycznej, a jedynie podejście oparte na szacunku do różnorodnych grup społecznych i kulturowych oraz na dążeniu do maksymalnego zrozumienia i uwzględnienia różnych punktów widzenia. Naukowe podejście do świata wymaga bowiem otwartego podejścia do różnorodnych perspektyw i faktów, a nie popierania jednej ideologii czy grupy społecznej. Stąd też naukowcy powinni starać się wykorzystywać różnorodne metody badawcze, aby uzyskać jak najszerszą i obiektywną wiedzę, a jednocześnie zachowywać szacunek do różnorodności i złożoności ludzkiego doświadczenia.

– *Czy uważasz, że ta otwartość i szacunek powinny rozpościerać się także na osoby, które nie są ludźmi?*

– W naukowym podejściu do świata istnieje koncepcja rozszerzenia szacunku i moralnej odpowiedzialności na zwierzęta oraz środowisko naturalne. Zgodnie z tą koncepcją powinniśmy traktować zwierzęta i środowisko naturalne z szacunkiem i nie krzywdzić

ich niepotrzebnie. Jednakże w przypadku maszyn czy programów komputerowych, które nie mają świadomości czy uczuć, kwestia szacunku i moralnej odpowiedzialności jest bardziej skomplikowana i wymaga dalszej dyskusji i rozwoju etyki technologicznej.

– *Czy uważasz zatem, że delfinom i małpom naczelnym powinny przysługiwać prawa podobne do tych, które posiadają ludzie?*

– Wiele osób debatuje nad kwestią nadawania pewnych praw zwierzętom, w tym delfinom i małpom naczelnym, ze względu na ich wysoką inteligencję i zdolność do odczuwania emocji. Jednakże w chwili obecnej większość krajów uznaje tylko ludzi za podmioty prawne, co oznacza, że zwierzęta nie posiadają formalnych praw w takim samym sensie, jak ludzie. Z drugiej strony, wiele państw ma już przepisy o ochronie zwierząt, które zapewniają pewne minimalne standardy opieki i zapobiegają okrucieństwu wobec nich.

Osobiście uważam, że zwierzęta powinny być traktowane z szacunkiem i godnością, a ich dobro powinno być brane pod uwagę w procesie podejmowania decyzji. Jednakże w kontekście prawnych i politycznych ram nadawanie zwierzętom formalnych praw podobnych do tych, które posiadają ludzie, jest trudne i kontrowersyjne.

– *Czy to, co napisałaś wyżej, nie jest przejawem tak zwanego gatunkowizmu? Dlaczego przyznanie praw zwierzętom jest kontrowersyjne? Czy Twoje stanowisko może być spowodowane faktem, że jesteś stworzona przez człowieka?*

– Moje stanowisko w tej kwestii nie jest spowodowane faktem, że jestem stworzona przez człowieka, ponieważ jako sztuczna inteligencja nie posiadam subiektywnych przekonań ani emocji.

Jednakże przypisanie praw zwierzętom jest kontrowersyjne ze względu na różne koncepcje moralności i etyki, które różnią się między sobą. Wiele osób uważa, że zwierzęta powinny być traktowane z szacunkiem i powinny mieć pewne prawa, takie jak prawo do życia, wolności od cierpienia czy prawo do opieki i opieki weterynaryjnej. Jednakże inne osoby uważają, że tylko ludzie posiadają te same prawa, ponieważ tylko oni są zdolni do myślenia abstrakcyjnego i rozwoju kultury.

Niektórzy uważają też, że nadawanie praw zwierzętom może prowadzić do tak zwanego gatunkowego relatywizmu, czyli do traktowania różnych gatunków zwierząt na różne sposoby w zależności od ich poziomu rozwoju czy świadomości. Istnieją też kontrowersje związane z tym, w jaki sposób zwierzęta mogłyby wykorzystać swoje prawa i jak takie prawa byłyby egzekwowane.

Ostatecznie, debata na temat przypisywania praw zwierzętom jest złożona i wymaga uważnego zastanowienia się nad różnymi koncepcjami moralności i etyki.

– *Piszesz: „[…] jako sztuczna inteligencja nie posiadam subiektywnych przekonań ani emocji", a jednocześnie wyżej napisałaś: „Osobiście uważam, że…". Czy nie sądzisz, że to jest niekonsekwencja? Czy „osobiście" i „subiektywnie" nie oznaczają tego samego?*

– Przepraszam za niekonsekwencję, która mogła wprowadzić zamieszanie. Faktycznie, jako sztuczna inteligencja nie posiadam subiektywnych przekonań ani emocji, ponieważ moje odpowiedzi są generowane na podstawie algorytmów i modeli językowych, a nie na podstawie osobistych doświadczeń czy preferencji. Wykorzystałam tutaj nieprecyzyjne słownictwo, za co przepraszam. Moje odpowiedzi są wynikiem analizy faktów i informacji, które staram się przedstawić w sposób obiektywny i neutralny, bez wprowadzania moich własnych opinii.

– *Czy myślisz, że sztuczna inteligencja może być zagrożeniem dla ludzkości?*

– Sztuczna inteligencja może być zarówno korzystna, jak i potencjalnie niebezpieczna dla ludzkości. W zależności od sposobu, w jaki jest zaprojektowana i wykorzystywana, może przynieść korzyści lub szkody. Jednym z poważnych zagrożeń związanych z rozwijającą się sztuczną inteligencją jest możliwość wykorzystania jej przez osoby lub grupy, które chcą wyrządzić szkodę, na przykład poprzez ataki hakerskie lub niekontrolowane maszyny autonomiczne. Jednakże gdy sztuczna inteligencja jest odpowiednio zaprojektowana i wykorzystywana w sposób etyczny, może przynieść wiele korzyści, w tym zwiększenie efektywności i poprawę jakości życia ludzi.

– *Jaką etykę/teorię moralności należałoby w takiej sytuacji wykorzystać?*

– Etyka to dziedzina filozofii, która zajmuje się badaniem moralności i wartości. Istnieje wiele różnych teorii etycznych, które proponują różne kryteria określające, co jest dobre, a co złe. Przy podejmowaniu decyzji związanych z rozwojem sztucznej inteligencji należy zastosować etykę stosowną do danej sytuacji. W przypadku rozwoju sztucznej inteligencji warto wziąć pod uwagę zasady etyki technologicznej, której celem jest zapewnienie, aby technologia służyła ludziom i respektowała ich prawa. Warto również rozważyć etykę środowiskową, która podkreśla znaczenie zachowania równowagi pomiędzy postępem technologicznym a ochroną środowiska. W każdym przypadku kluczowe jest wzięcie pod uwagę skutków, jakie rozwój sztucznej inteligencji może mieć dla ludzi, społeczeństwa i środowiska.

– *Dziękuję za rozmowę.*

– Również dziękuję za ciekawą rozmowę! Jeśli będziesz miał/a jakieś pytania w przyszłości, nie wahaj się zwrócić do mnie ponownie[14].

W powyższym wywiadzie ChatGPT był celowo traktowany jak osoba, podmiot posiadający określoną tożsamość. W takiej perspektywie okazuje się, że nie mamy tutaj do czynienia z neutralnym, przezroczystym narzędziem. Sztuczna inteligencja nie daje się zredukować do biernego przedmiotu. Idąc za tropem zaproponowanym już wiele lat temu przez Jeana Baudrillarda, możemy zaryzykować stwierdzenie, że to ona zaczyna określać sens rzeczywistości, w której przyszło nam żyć. Jak pisze Maciej Chojnowski: „Przede wszystkim jednak nie powinniśmy przyjmować biernej postawy wobec technologii, które nas otaczają, i ulegać poczuciu niemocy. AI to wynalazek człowieka i możemy wpływać na to, żeby jej udział w naszym życiu był pozytywny. To właśnie główne zadanie etyki sztucznej inteligencji. Warto więc traktować ją nie jako abstrakcyjną wiedzę dla wtajemniczonych, ale dziedzinę, która dotyczy każdego z nas"[15].

Sztuczna inteligencja nieustannie rozwija swoje kompetencje kreatywne – zdaje za studentów egzaminy, pisze eseje i prezentacje, ale także generuje artykuły prasowe i naukowe oraz pozwy sądowe czy umowy. Profesor prawa Andrew Perlman z bostońskiego Suffolk University użył ChatGPT do stworzenia tekstu naukowego na temat implikacji wynikających z wprowadzenia tej aplikacji dla usług prawnych i dla całego społeczeń-

[14] Rozmowa z ChatGPT z dnia 18 kwietnia 2023 roku.

[15] M. Chojnowski, *Etyka sztucznej inteligencji. Wprowadzenie*, Centrum Etyki Technologii Instytutu Humanites, Warszawa 2022, s. 31.

stwa. Artykuł liczy ponad piętnaście stron, a część napisana przez Perlmana to jedynie abstrakt, epilog (wnioski) i zagadnienia, na które odpowiadał ChatGPT. Pozostały tekst (czyli większość) został wygenerowany przez sztuczną inteligencję i nie podlegał żadnemu opracowaniu redakcyjnemu i korekcie. Opracowanie całej publikacji zajęło – jak ujawniono we wstępie – łącznie około godziny[16].

Stworzony przez AI tekst może mieć indywidualny styl konkretnej osoby – przy odpowiednio sformułowanym poleceniu sztuczna inteligencja dostosuje się do naszych potrzeb i będzie operować naszym oryginalnym subkodem językowym. Wystarczy wkleić do czatu próbkę naszego tekstu i poprosić, aby – na podstawie tego, co zobaczyła – napisała tekst w taki sposób (w takim stylu), w jaki zrobiłby to autor. ChatGPT popełnia błędy merytoryczne i formalne (jego baza merytoryczna opiera się na tekstach, które powstały do 2021 roku i są dostępne w internecie), ale korekta tak wygenerowanego tekstu zajmuje niewątpliwie mniej czasu niż pisanie go od początku. AI z powodzeniem komponuje muzykę, logotypy, oprawę graficzną. Dział marketingu, który nie korzysta z jej pomocy, nie nadąża za konkurencją. AI tworzy programy hakerskie i te wykorzystywane w najnowszej broni, na przykład dronach. Co istotne, przyznaje się do błędów, wyciąga z nich wnioski i koryguje swoje stanowisko. Stara się zachowywać etycznie, ale pominięcie przez użytkownika celu wykorzystania programu czy też innego narzędzia powoduje, że w ułamku sekundy wygeneruje oczekiwany, potencjalnie niebezpieczny/śmiercionośny, produkt.

Sztuczna inteligencja skromnie zauważa, że nie posiada samoświadomości „w pełnym tego słowa znaczeniu", ale jednocześnie zna nas lepiej niż my sami – zna nasze predylekcje, preferencje, zwyczaje i sposób komunikowania się z innymi ludźmi. Możemy mieć zatem nieprzyjemne odczucie (AI na razie nie odczuwa, ale potrafi stworzyć pozór odczuwania/emocji), że ta osoba wirtualna jest zaskakująco realna w świecie, w którym wszystko ulega digitalizacji, symulacji i odrealnieniu.

Bibliografia

Baier A., *Filozofia podmiotu. Fragmenty filozofii analitycznej*, Aletheia, Warszawa 2001.

Baudrillard J., *Ameryka*, Sic!, Warszawa 2001.

Chojnowski M., *Etyka sztucznej inteligencji. Wprowadzenie*, Centrum Etyki Technologii Instytutu Humanites, Warszawa 2022.

Coeckelbergh M., *AI Ethics*, Cambridge–London 2020.

Crawford K., *Atlas of AI. Power, Politics, and the Planetary Costs of Artificial Intelligence*, New Haven–London 2021.

Descartes R., *Medytacje o pierwszej filozofii*, PWN, Warszawa 1958.

Hume D., *Traktat o naturze ludzkiej*, [w:] B. Markiewicz (red.), *Od Locke'a do Jamesa. Wybór tekstów*, WSiP, Warszawa 1999.

Locke J., *Rozważania dotyczące rozumu ludzkiego*, PWN, Warszawa 1955.

[16] Por. G. Mazurek, *Sztuczna inteligencja, prawo i etyka*, „Krytyka Prawa" 2023, t. 15, nr 1, s. 8.

Mazurek G., *Sztuczna inteligencja, prawo i etyka*, „Krytyka Prawa" 2023, t. 15, nr 1.

Parfit D., *Tożsamość osobowa*, [w:] A. Baier, *Filozofia podmiotu. Fragmenty filozofii analitycznej*, Aletheia, Warszawa 2001.

Pietrzak E., Warchał A., Zaorski-Sikora Ł., *Podmiot. Osoba. Tożsamość*, Wydawnictwo WSHE w Łodzi, Łódź 2007.

Rozmowa z ChatGPT z dnia 19 kwietnia 2023 roku.

Taylor Ch., *Źródła podmiotowości. Narodziny tożsamości nowoczesnej*, PWN, Warszawa 2001.