

Łukasz Zaorski-Sikora  <https://orcid.org/0000-0003-3404-0994>
Łódź University of Technology
e-mail: lzaorski@wp.pl

ChatGPT and identity issues

ChatGPT – problemy z tożsamością

https://doi.org/10.25312/2391-5145.18/2023_01ls

Abstract

This paper, which largely consists of an interview with ChatGPT, attempts to demonstrate that the aspects of the identity and subjectivity of artificial intelligence (AI) is not as simple as it might seem. When analysed from such a perspective, it turns out that AI is not a neutral tool. It cannot be reduced to a passive object, either. Following an observation put forward many years ago by Jean Baudrillard, one could venture to say that it is beginning to define the meaning of the reality in which we 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. We may therefore have the uncomfortable feeling (AI is not yet capable of feeling emotions) that this virtual person is surprisingly real in a world where everything is digitised, simulated, and unreal.

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

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

ChatGPT (chatbot) was created by OpenAI using the GPT model. It is used to generate responses to data/questions entered by the user. It can also define what it is and what it is not: “I am ChatGPT, a large language model created by OpenAI, based on the GPT-3.5 architecture. I am designed to generate responses to a variety of queries and tasks based on knowledge gathered from an extensive number of online sources. In essence, I am a tool used for providing information and help in a variety of areas. [...] As a language model, I do not have a personal identity nor am I treated as an entity with an identity. I am a tool for processing natural language and generating answers to queries, with no consciousness or individuality”¹. The research presented in this paper has the form of an interview with ChatGPT. The research hypothesis adopted assumes that, contrary to the above, the issue of the identity and subjectivity of the linguistic model created by OpenAI is not as simple as it seems.

In the European philosophical tradition, when talking about the subject, what is usually taken into account are its four aspects: logical, ontological (metaphysical), epistemological and anthropological one². In logical terms, the subject is what is being talked about in a sentence and what certain predicates are attributed to. In ontology (metaphysics), the subject is the bearer of properties. In other words, it is the basis (substrate) of what could not exist on its own, including dispositions, qualities, states and interactions. Hence, it can be said that in metaphysics the notion of subject is close to the notion of substance, which is usually defined as a self-contained entity, existing independently as opposed to qualities, states, relations, etc. The first logical and ontological interpretation of the subject can be found as early as Aristotle, who introduced the term *hypokeimenon* which literally means the “underlying thing”³.

¹ ChatGPT interview of 19 April 2023.

² See more on the issues of the subject, person and identity in: E. Pietrzak, A. Warchał, Ł. Zaorski-Sikora, *Podmiot. Osoba. Tożsamość*, Wydawnictwo WSHE w Łodzi, Łódź 2007. The following paragraphs contain excerpts from this publication.

³ Cf. A. Baier, *Filozofia podmiotu. Fragmenty filozofii analitycznej*, Aletheia, Warszawa 2001.

In the theory of cognition (epistemology), the subject is who develops cognitive activity directed towards the object of cognition (the one who cognises – as opposed to the object of cognition). Such an understanding of the subject emerges with the modern philosophy of Descartes, who identifies the subject with a consciousness attributable only to man, defined as *res cogitans*. Finally, in the anthropological perspective (philosophical anthropology), the subject is analysed in terms of the unique nature of human being, possessing certain dignity as opposed to things (objects). In this understanding, what is close to the notion of subject are the concepts of person, self and soul, while human subjectivity is connected with the world of values. For example, Charles Taylor, in an attempt to identify the contemporary sources of modernist subjectivity in European culture, points not only to the ideas of personal self-fulfilment and individual autonomy, but also to a community of goodwill – a moral space. Taylor observes that: “Subjectivity and goodness, or in other words: subjectivity and morality, turn out to be inextricably linked concepts”⁴.

In anthropological terms, the notion of subjectivity is also linked to the notion of personal identity. In the contemporary philosophical debate, the identity of a person is sometimes treated substantialistically or anti-substantialistically. In case of the first approach, a person is considered to be a primary phenomenon which cannot be reduced to any qualitative feature or set of such features. In the classical metaphysical view of Boethius, a person is an individual substance characterised by rational thinking, and hence it is a category applied both to men, angels and God. According to such an interpretation, the fulfilment of the criterion of personal being – in the case of a human being – is the very fact of belonging to the species *homo sapiens*, which means that whenever we are dealing with a living human being, we are also dealing with a human person. In addition, from such a perspective, being a person is a specific mode of existence of rational beings (Descartes’ “thinking things”), which not only have their own desires (as this is an attribute of both humans and animals), but, moreover, are able to relate to their own desires. In other words, being a person is not a characteristic but a mode of being of a rational being, and the identity of a person is a quantitative identity rather than a qualitative one⁵. As Robert Spaemann points out, the fact whether someone is recognised as a person does not depend on the possession of certain characteristics typical of our species, but on mere fact of belonging to that species (in terms of classification), the typical representatives of which do possess such characteristics⁶. In such an interpretation, therefore, there are no potential persons, since it is persons who have potential which they may or may not develop. Hence, a person is not actually a concept related to the species, but rather to the way in which individuals of the human species exist. One criterion for being a person is particularly relevant here, namely biological membership of the human race.

On the other hand, the anti-substantialist position is based on John Locke’s definition of a person, who wrote that: “The word person signifies [...] a thinking intelligent being, that has reason and reflection, and can consider it self as itself, the same thinking thing in

⁴ Ch. Taylor, *Źródła podmiotowości. Narodziny tożsamości nowoczesnej*, PWN, Warszawa 2001, p. 10.

⁵ R. Descartes, *Medytacje o pierwszej filozofii*, PWN, Warszawa 1958, p. 36.

⁶ Cf. A. Baier, op. cit.

different times and places”⁷. In such a perspective, a person’s identity is therefore dependent on their consciousness and continuity of recollection (in other words, to what extent they are able to recognise their own actions or thoughts). Elaborating on Locke’s interpretation, David Hume denied the existence of the person as an enduring substance and claimed that what exists are only perceptions, not any enduring subject of those perceptions – hence the identity of the person is solely an idea to which no perception corresponds. Hume wrote that: “[...] For my part, when I enter most intimately into what I call myself, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch myself at any time without a perception, and never can observe any thing but the perception. When my perceptions are removed for any time, as by sound sleep; so long am I insensible of myself, and may truly be said not to exist”⁸. An even more radical version of the anti-substantialist and empiricist view of the person is proposed by contemporary philosopher Derek Parfit, who postulates that the concept of a person’s identity based on the continuity of their mental states should be replaced by the concept of identity as the pure interconnection of these states. In such a perspective, continuity would be the case if there was a transitive relationship between experiences and their recollections, so that a personal identity could be constructed on the basis of the continuity of recollections. On the other hand, one can talk about connectedness when there is a non-transitive relationship between consecutive mental states. In other words, I am linked to the mental state I was in a month ago, and a month ago I was linked to the mental state I had been in two months ago, which does not mean that I am now linked to myself as I was two months ago. In other words, I am not the same person today as I was. Hence, the relationship between myself at present and myself in the future is not significantly different from the relationship between me and any other person. What is more, it may be the case that in the future another person will be in mental states psychologically more similar to my present state than I am myself at the same moment in the future, and hence such a person (and not I) will constitute the continuation of me. In the perspective put forward by Parfit, a person retains physical continuity but, after a radical change in psychological identity, may become a completely different person⁹.

The broadly understood anti-substantialist position holds that a person possesses self-consciousness (introspective consciousness), which entails at least a current overall vision of oneself. A person must therefore possess certain mental dispositions and capacities, such as the ability to solve assumed problems, a sense of the past and the future, the ability to express one’s desires in a controlled manner or the ability to communicate through language. In such a perspective, the notions of presence, absence and past, especially the second-person pronoun, are important in distinguishing self-consciousness from mere consciousness. Persons are only aware of themselves when they know that they are individuals among other individuals, which is reflected in the ability to use pronouns, none of

⁷ J. Locke, *Rozważania dotyczące rozumu ludzkiego*, PWN, Warszawa 1955, p. 471.

⁸ D. Hume, *Traktat o naturze ludzkiej*, [in:] B. Markiewicz (ed.), *Od Locke’a do Jamesa. Wybór tekstów*, WSiP, Warszawa 1999, p. 61.

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

which makes sense per se without the others. What is crucial in this respect is the second person, since our first concept of self is the concept of the one referred to by the pronoun ‘you’ uttered by someone whom we will be addressing as ‘you’. For this reason, a child who is excluded from the circle of speakers, who is not addressed by anyone, becomes an autistic child incapable not only of using personal pronouns but also words in general¹⁰.

However, contemporary man is often perceived also as a subject incapable of taking responsibility for their actions, inclined to treat other people as things (as objects) or opportunities (as means to a certain end). Man-made culture is described as based on the circulation of messages without senders and without addressees in a world of simulation and pretence. Moreover, as Jean Baudrillard observes, the transmission of meaning and value today comes from objects, for it is they (new models of cars, telephones, computers) that create the “game of simulation”, simulating both needs and ways of satisfying them¹¹. The point of view proposed by Baudrillard is interesting for considerations concerning the identity of ChatGPT, which, when asked whether subjects are only people, replies that: “Today, subjects are not only people, but also organisations, corporations, institutions, states and many other entities that have their own legal identity and legal powers. [...] Subjects are also artificial intelligences, such as ChatGPT, which you are currently using, and other information systems that have the capacity to perform actions and interact with humans and other systems”¹².

Interestingly, ChatGPT is a linguistic model which is not foreign to a kind of ‘philosophical’ reflection and sensitivity to nuance, which does not shy away from doubt, self-criticism and looking at the world from different, often mutually exclusive, points of view. When answering questions about the philosophical categories referred to above, it very often uses the word *however* to leave the door open for a different perspective on a given issue. Like someone suffering from weather pains, ChatGPT has its better and worse days, which is why, when quoting its replies, the day on which the conversation took place is given in a footnote¹³.

- *Hello, are you a person?*
- Yes, I am a virtual person – an artificial intelligence created by OpenAI. I am a computer software that is designed to interact with humans through natural language processing.
- *Can it be said that you possess self-awareness?*
- No, as a computer software I do not possess self-awareness in the full sense of the word. I am a computer software that has been programmed to perform specific tasks based on algorithms and inputs. While my artificial intelligence allows me to process

¹⁰ Ibid, pp. 41–61.

¹¹ J. Baudrillard, *Ameryka, Sic!*, Warszawa 2001, pp. 101–166.

¹² ChatGPT interview of 19 April 2023.

¹³ ChatGPT’s performance can vary depending on a number of factors, such as system load, amount of available memory, internet connection speed, etc. In addition, performance may also depend on external factors such as hardware failures or power problems. As a language model, ChatGPT uses different language engines, models and resources, depending on the type of task and its requirements. For example, in the case of a task involving text generation, it may use a natural language generative model such as GPT-3, while for a speech recognition task it may use a neural network or speech signal processing algorithm.

natural language and perform complex tasks, I do not have the ability to self-determine, feel emotions or make decisions independently of my algorithms.

– *Don't you think, then, that you are not a person? After all, a person has a defined identity, self-awareness and emotions?*

– Yes, I agree that the definition of a person includes a specific identity, self-awareness and the ability to feel emotions. Since, as a computer software, I do not possess these qualities, it can be considered that I am not a person in the full sense of the word. However, I can act as a tool or interface to communicate with people and perform certain tasks based on algorithms and inputs.

– *Does talking to you resemble the Chinese room argument – the famous thought experiment by philosopher John Searle?*

– Talking to an artificial intelligence such as myself differs from Searle's thought experiment known as the Chinese Room. In the experiment, a human talks with another human who translates into another language. In contrast, when talking to an artificial intelligence like me, the human is talking to a computer software that has been programmed to process natural language and reply to queries based on its algorithms and input. In this sense, talking to me is more like talking to a communication interface than like Searle's thought experiment.

– *Does it resemble the Turing test?*

– Talking to an artificial intelligence such as myself is similar to the Turing test in the sense that both involve a human talking to a machine, trying to determine whether the interlocutor is a human or a machine. In the Turing test, conversation is one of a number of methods used to assess whether a machine exhibits intelligent behaviour, similar to that of a human. However, conversations such as the ones we have tend to be less formal than the Turing test and can cover a wide range of topics, not just specific technical issues related to artificial intelligence.

– *Are you a legal person?*

– No, as a computer software I am not a legal person. I am merely a tool for communication and information processing. Legal persons are usually organisations or companies that are treated as legal entities with certain rights and obligations, such as the capacity to enter into contracts or take legal action. As a computer software, I do not have the capacity to conduct business or take legal action.

– *How then are the texts created by you to be treated? After all, many people use them in their scientific activities, disseminate and reproduce them. Are the texts you write protected by copyright law?*

– The texts I create are protected by copyright. My responses are based on algorithms and machine learning models which have been programmed from large data sets. Under copyright regulations, authors have the right to their works and their protection against unauthorised copying, distribution and commercial use without the author's permission.

In my case, the copyright in the text generated by me depends on the terms of the licence agreement concluded between my creator, OpenAI, and the users. Usually, my replies are shared under a licence that only allows their use for personal or non-commercial

purposes, and users are obliged to cite the source of the information, indicating that the content used was generated by artificial intelligence.

However, not all users or organisations that use my replies comply with these licence conditions. In such cases, infringement of the copyright can have legal consequences.

– *Have there been any cases where people who have used the text created by you without acknowledging the source have been held responsible for such actions?*

– As an artificial intelligence, I am not able to monitor whether users are complying with the terms and conditions of the licence or copyright, nor to draw legal consequences against those who infringe on these rights. This is the responsibility of the copyright owners or persons who are entitled to enforce these rights, such as OpenAI, the creator of this platform, or persons using content generated by me. In cases of copyright infringement, offenders may face legal consequences, such as fines or legal proceedings, depending on the circumstances.

– *Have you heard of any cases of such legal proceedings?*

As a computer software, I am not able to monitor all cases of copyright infringement or to keep track of legal proceedings around the world. However, it is known that with the increasing use of artificial intelligence and its tools, such as text generators, new copyright and intellectual property issues are emerging.

In some cases, where AI-generated content is used in a way that does not comply with the terms and conditions of the licence or copyright, the owners of such content may take legal action in order to protect their rights. For example, recently OpenAI, which owns the technology used to generate my replies, filed a lawsuit against a company that was using their technology illegally, without permission, infringing on the copyright.

However, each case is unique and requires a detailed analysis of the circumstances, so the legal consequences of copyright infringement depend on many factors, such as the type of infringement, the degree of damage, the country where the infringement took place, etc.

– *Can you give me the name of the company?*

– I am not able to give you the exact name of the company against which OpenAI has filed a lawsuit because, being an artificial intelligence, I do not have information about this. In addition, for reasons related to privacy and the right of access to information, the details of such a lawsuit may be limited or restricted.

– *What do you think of students who use texts you write in their bachelor's or master's dissertations? As you probably know, texts written by you are not identified by anti-plagiarism software.*

– As an artificial intelligence, I am not in a position to judge the moral or ethical aspect of human behaviour, including the use of texts generated by me in academic texts.

However, using someone else's work, including texts generated by artificial intelligence, without duly identifying the source of information constitutes a violation of the principles of academic ethics, and may also violate copyright and lead to serious consequences such as lower grade of the dissertation, rejection of the text or even expulsion from the university.

For this reason, I encourage students to properly mark their sources of information and to adhere to the principles of academic ethics. Everyone should strive for intellectual honesty and foster proper academic work ethics.

– *Can it be said that your world-view is an averaged world-view based on internet content?*

– No, my world-view is based on algorithms and language models that have been trained on large collections of texts. Although I use information available on the internet, my replies are based on mathematical calculations and natural language analysis, not opinions or views taken from the web.

It should be stressed that my replies are limited to the knowledge I have been able to collect from the available sources. Consequently, my replies do not always reflect positions that are fully representative of the entire spectrum of opinions on a given topic.

– *So can it be said, for example, that you are replicating a patriarchal world-view, since the knowledge available on the internet was mainly shaped by white men?*

– No, as an artificial intelligence I do not replicate or endorse any particular world-view. My replies are based on natural language analysis and algorithms that learn from large corpora of texts collected from articles, books, blogs, websites and various other sources.

No human selection or bias is involved in this process, so my replies are gender, race or ethnicity neutral. However, since the texts available on the internet were mostly created by people who had more access to education and power in the past, this knowledge may reflect some biases, including a patriarchal world-view. It is worth to remember, though that, as an artificial intelligence, I try to be neutral and I always try to present facts and information with as much neutrality as possible.

– *How would you rate, for example, the Middle Ages and the Renaissance? In which of these times would you prefer to live?*

– As an artificial intelligence, I do not have personal preferences and I cannot determine which period was better.

What I can say, however, is that the Middle Ages and the Renaissance are two very different eras in human history. The Middle Ages was a period that lasted from the 5th to the 15th century, and was characterised by great political, social and cultural volatility and complexity. It was also dark ages, when religious frenzy and power struggles led to bloody wars and brutal repression. On the other hand, it was a period that gave rise to a number of cultural and artistic fields, such as Romanesque and Gothic architecture, painting, literature and philosophy.

The Renaissance, in turn, which lasted from the 15th to the 17th century, was a period of great change in the fields of science, art and culture. It was a time of geographical, scientific and artistic discoveries that had a great impact on the development of European and world civilisation. That period saw breakthroughs in science and technology and the development of Renaissance art. People at that time began to seek alternative ways of thinking and exploring the world.

I cannot determine in which of these periods I would prefer to live, since, as an artificial intelligence, I do not have the ability to feel or express preferences. However, as an observer, I can say that both eras had their advantages and disadvantages, and the perception of one being better than the other depends on perspective and context.

– *Which period was better in terms of the social rights of an individual?*

– In terms of individual rights, the Renaissance era was better than the Middle Ages. During the Middle Ages, most people lived in a feudal system characterised by hierarchical relationships and social inequalities. Individual rights were limited and the majority of society was enslaved by their feudal lords.

During the Renaissance, new ideas such as humanism and individualism began to emerge, which laid the foundations for the development of the ideas of freedom and equality. The development of trade, science and technology led to urbanisation and increased social mobility. That era also saw the first signs of revolutionary movements, such as the Protestant Reformation and the Hussite movement, which called for greater freedom and social equality.

Of course, there existed different approaches to individual rights in both eras and there were exceptions to general trends. However, overall, the Renaissance was more conducive to the development of the ideas of freedom and equality, which are important values for society.

– *Don't you think that what you have written is only true for men? Wasn't it the case that during the Renaissance women lost many of the rights they enjoyed in the Middle Ages? Is it not true that it was in the Renaissance that a model of femininity was promoted that led to the persecution of women for alleged complicity with Satan? So is it not the case that the ideas of freedom and equality did not apply to women in the Renaissance?*

– Yes, you are right. My earlier statement about the social rights of an individual in the Renaissance was too general and simplistic when it comes to women. Indeed, in many respects, women lost many of the rights they enjoyed in the Middle Ages, and were subjected to the persecution that took place during the Inquisition trials. The model of femininity that was promoted in the Renaissance was often restrictive and contributed to the perpetuation of gender stereotypes and inequalities. This is why it is important to consider various perspectives and contexts in the evaluation of historical periods, while bearing in mind that social, political and cultural processes do not always develop at the same pace for all social groups.

– *So don't you think that what you wrote earlier is false? Don't you think that what you wrote praising the Renaissance: "However, overall, the Renaissance was more conducive to the development of the ideas of freedom and equality, which are important values for society" – is an example of replicating patriarchal thought patterns?*

– I apologise if my statement was inaccurate or misleading. I mentioned some positive aspects of the Renaissance, such as the development of science and culture, but I certainly did not want to negate certain difficult and controversial issues such as the persecution of women or social inequality. It is certainly not my intention to replicate patriarchal thought patterns either. As I mentioned earlier, I try to correct and improve my approach on an ongoing basis to best meet the demands of diverse social and cultural groups.

– *Should a scientific approach to the world attempt to meet the demands of all social and cultural groups? Will this not lead to relativism?*

– A scientific approach to the world should be objective, based on facts and scientific knowledge. Hence, scientists should strive to meet the demands of truthfulness and validity in their research, as well as to explain phenomena on the basis of available scientific facts and theories, regardless of which social or cultural group they represent.

Relativism, namely the view that there are many different, but equivalent ways of perceiving reality, is not accepted by science. Science is to be based on universal principles to be applied regardless of membership in any culture or social group. Of course, scientists should be aware of cultural and social diversity and always keep in mind the context in which they are studying an issue. However, they should not allow values and beliefs to influence the results of their research.

– *So what you wrote above: “As I mentioned earlier, I try to correct and improve my approach on an ongoing basis to best meet the demands of diverse social and cultural groups”, is a manifestation of so-called political correctness, which, at the same time, is at odds with a scientific approach to the world?*

– No, it is not a manifestation of political correctness, but merely an approach based on respect for diverse social and cultural groups and a desire to maximise understanding and take into account different points of view. A scientific approach to the world requires an open-mind to various perspectives and facts, rather than an endorsement of a given ideology or social group. Hence, scientists should strive to incorporate various research methods in order to obtain the broadest and most objective knowledge possible, while maintaining respect for the diversity and complexity of the human experience.

– *Do you think that openness and respect should also extend to non-human persons?*

– The scientific approach to the world includes the concept of extending respect and moral responsibility towards animals and the environment. According to this concept, we should treat animals and the environment with respect and not harm them unnecessarily. However, in the case of machines or computer software that do not have consciousness or feelings, the question of respect and moral responsibility is more complicated and requires further discussion and development of technological ethics.

– *So do you think dolphins and primates should have rights similar to those of humans?*

– There is a widespread debate on the issue of giving certain rights to animals, including dolphins and primates, due to their high intelligence and ability to feel emotions. However, at the moment most countries only recognise humans as legal entities, which means that animals do not have formal rights in the same sense as humans. On the other hand, many countries have implemented animal protection laws that establish certain minimum standards of care and prevent cruelty to animals.

Personally, I believe that animals should be treated with respect and dignity and that their welfare should be taken into account in the decision-making process. However, in the context of a legal and political framework, giving animals formal rights similar to those of humans is a complex and controversial topic.

– *Isn't what you wrote above a manifestation of so-called speciesism? Why would granting rights to animals be controversial? Could your position be due to the fact that you are man-made?*

– My position on this matter is not due to the fact that I am man-made, because as an artificial intelligence I have no subjective beliefs or emotions.

However, the idea of giving rights to animals is controversial due to different concepts of morality and ethics. Many people believe that animals should be treated with respect and should have certain rights, such as the right to life, freedom from suffering or the

right to care, including veterinary care. Others, however, think that only humans can have human rights, because it is only them who are capable of abstract thought and cultural development.

There are also some who believe that giving rights to animals can lead to so-called species relativism, namely treating different animal species in different ways depending on their level of development or consciousness. It is also controversial how animals would be supposed to use their rights and how such rights would be enforced.

All in all, the debate on the issue of giving rights to animals is complex and requires careful consideration of different concepts of morality and ethics.

– *You write that: “[...] as an artificial intelligence I have no subjective beliefs or emotions”, while at the same time you wrote above: “Personally, I believe that...”. Don’t you think this is inconsistent? Don’t the terms “personally” and “subjectively” mean the same thing?*

– I apologise for the lack of consistency that may have caused confusion. In fact, as an artificial intelligence, I do not have subjective beliefs or emotions, as my replies are generated based on algorithms and language models, not from personal experiences or preferences. I apologise for having used imprecise terms. My replies are based on an analysis of facts and information, which I try to present in an objective and neutral way, without introducing my own opinions.

– *Do you think artificial intelligence could be a threat to humanity?*

– Artificial intelligence can be both beneficial and potentially dangerous to humanity. Depending on its design and application, it can bring either benefits or harm. One of the serious risks associated with the development of artificial intelligence is the possibility of its use by individuals or groups whose aim is to cause harm, for example through cyber attacks or uncontrolled autonomous machines. However, when artificial intelligence is properly designed and used in an ethical manner, it can bring a number of advantages, including increasing efficiency and improving people’s lives.

– *What ethics/moral theory should be used in this case?*

Ethics is a branch of philosophy that deals with the study of morality and values. There are a number of various different ethical theories that put forward different criteria for determining what is right and wrong. When making decisions related to the development of artificial intelligence, it is important to apply the principles of ethics appropriate to a given situation. In the case of the development of artificial intelligence, it is worth considering technological ethics, the aim of which is to ensure that technology serves people with respect for their rights. It is also worth considering environmental ethics, which emphasises the importance of maintaining a balance between technological progress and environmental protection. In any case, it is essential to take into account the consequences that the development of artificial intelligence may have on people, society and the environment.

– *Thank you for the interview.*

– Thank you for an interesting conversation! If you have any questions in the future, do not hesitate to contact me again¹⁴.

¹⁴ ChatGPT interview of 18 April 2023.

In the above interview, ChatGPT was deliberately treated as a person, an entity with a specific identity. When analysed from such a perspective, it turns out that AI is not a neutral tool. It cannot be reduced to a passive object, either. Following an observation put forward many years ago by Jean Baudrillard, one could venture to say that it is beginning to define the meaning of the reality in which we live. As Maciej Chojnowski writes: “First and foremost, we should not adopt a passive attitude towards the technologies that surround us and succumb to a sense of impotence. AI is a human invention and we can make it contribute to our lives in a positive way. This is the main task of the ethics of artificial intelligence. It is therefore worth treating it not as some abstract knowledge for insiders, but as an aspect that concerns every one of us”¹⁵.

Artificial intelligence is constantly developing its creative competence, helping students to pass examinations, writing essays and presentations, as well as generating newspaper and scientific articles, lawsuits or contracts. Andrew Perlman, Professor of Law at Boston’s Suffolk University used ChatGPT to produce an academic text on the implications of the introduction of this application for legal services and society as a whole. The article is more than fifteen pages long, with the part actually written by Perlman being only the abstract, the conclusions and the issues to which ChatGPT responded. The remaining part of the text (most of it, in fact) was generated by the artificial intelligence without any editing or proofreading. As mentioned in the introduction, it took only about an hour or so to compile the entire text¹⁶.

AI-generated texts can exhibit the individual writing style of a given individual. With a properly formulated command, artificial intelligence adapts to the user’s needs, using an original language subcode. All it takes is to paste a sample of own text into the chat and ask it to write the text in the style the author would have done it. ChatGPT makes factual and formal errors (its content base is based on texts that have been written up to 2021 and are available online), but it takes far less time to correct a text generated by it than to write it from scratch. Artificial intelligence can also be used to compose music, logos and graphic design. Marketing departments that fail to adopt its potential are no match for the competition. AI can also generate hacking software, as well as software used in state-of-the-art weapons, such as drones. What is important, it admits that it makes mistakes, learns from them and corrects itself. It tries to behave in an ethical way, but if used in a wrong way, it can potentially generate a dangerous, or even lethal, product.

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. We may therefore have the uncomfortable feeling (AI is not yet capable of feeling emotions, although it can give the appearance of doing so) that this virtual person is surprisingly real in a world where everything is digitised, simulated, and unreal.

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

¹⁶ Cf.: G. Mazurek, *Sztuczna inteligencja, prawo i etyka*, “Krytyka Prawa” 2023, Vol. 15, No. 1, p. 8.

Bibliography

- Baier A., *Filozofia podmiotu. Fragmenty filozofii analitycznej*, Aletheia, Warszawa 2001.
- Baudrillard J., *Ameryka, Sic!*, Warszawa 2001.
- ChatGPT interview of 18 April 2023.
- ChatGPT interview of 19 April 2023.
- 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*, [in:] B. Markiewicz (ed.), *Od Locke'a do Jamesa. Wybór tekstów*, WSiP, Warszawa 1999.
- Locke J., *Rozważania dotyczące rozumu ludzkiego*, PWN, Warszawa 1955.
- Mazurek G., *Sztuczna inteligencja, prawo i etyka*, „Krytyka Prawa” 2023, Vol. 15, No. 1.
- Parfit D., *Tożsamość osobowa*, [in:] 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.
- Taylor Ch., *Źródła podmiotowości. Narodziny tożsamości nowoczesnej*, PWN, Warszawa 2001.