ŁUKASZ ABRAMOWICZ

ONE DOES NOT SIMPLY CREATE AN AI ISSUE. ON THE PSEUDO-PROBLEMATIC NATURE OF AI ISSUE

ABSTRACT. Łukasz Abramowicz, *One does not simply create an AI issue. On the pseudo-problematic nature of AI issue*, edited by Sławomir Leciejewski, „Człowiek i Społeczeństwo" vol. LVIII: *Społeczny wymiar rewolucji informatycznej* [The social dimension of the information technology revolution], Poznań 2024, pp. 131–150, Adam Mickiewicz University. ISSN 0239-3271, https://doi.org/10.14746/cis.2024.58.7.

The debate surrounding the topic of Artificial Intelligence (AI), and its different meanings, seems to be ever-growing. This paper aims to deconstruct the seemingly problematic nature of the AI debate, revealing layers of ambiguity and misperceptions that contribute to a pseudo-problematic narrative. Through a review of existing literature, ethical frameworks, and public discourse, this essay identifies key areas where misconceptions, hyperbole, and exaggerated fears have overshadowed the genuine concerns associated with AI development and deployment. To identify these issues I propose three general criteria that are based on Popper's and Ayer's work and adjusted to my needs. The subsequent sections categorize AI issues into ontological, methodological, and logical-grammatical problems, aligning with Cackowski's typology. In addition, I introduce «» signs to distinguish behavioural descriptions from cognitive states, aiming to maintain clarity between external evidence and internal agent states. My conclusion is quite simple: the AI debate should be thoroughly revised, and we, as scholars, should define the concepts that lie at the bottom of AI by creating a universal terminology and agreeing upon it. This will give us the opportunity to conduct our debates reasonably and understandably for both scholars and the popular public.

Keywords: Artificial Intelligence, pseudo-problem, intelligence, consciousness

Łukasz Abramowicz, Adam Mickiewicz University in Poznań, Faculty of Psychology and Cognitive Science, ul. Szamarzewskiego 89AB, 60-568 Poznań, e-mail: abramowicz.lukasz18@gmail.com, https://orcid.org/0009-0009-8355-0924.

Introduction

In the literature, there are multiple discussions about Artificial Intelligence (AI) understood as a domain, an artefact, or the AI thesis. All the meta-theoretical considerations, on the three basic meanings of AI, are themselves referred to as the AI issue. The AI domain is, primarily, about the research of AI. Philosophers often argue about how to evaluate AI paradigms, whether the research of AI is scientifically useful, and if so, then how. In terms of AI as an artefact which is the product of AI research, one might think about how and why such an artefact would function and how one should treat such an artefact. Could it be a moral patient or perhaps even a moral agent? Ought it to have some rights and be held accountable for its 'actions'? As for the AI thesis, it is the claim that the creation of an AI artefact is possible. As shown in this overview, the range of these debates is very broad.

The purpose of this paper is to explore discussions surrounding AI, in search of their flaws. I could not find any recent sources that would thoroughly reconsider the ongoing debates, and I want to fill this gap by making certain general points as a starting point for reassessing the AI-related debates. Throughout the course of this article, I shall examine some selected most common AI issues and evaluate them according to the criteria of being a pseudo-problem. More specifically, I shall consider: the empirical character of AI issues, such as ethical considerations; the question of whether the AI thesis is sound; whether AI research might tell us something new about the world; and the conceptual confusion surrounding discussions of AI. Now, one might ponder why even bother with such a purpose. Well, it is very important to do science responsibly. Arguing about pseudo-problems is merely tilting at windmills and does not enhance our knowledge about the world, often hindering the development of science. Moreover, as we shall see, they generate many discussions that consume time and energy. Instead, by pointing out the pseudo-problems, we could try a different approach to the considered problems or focus on more fruitful topics that are worth our time and could develop our knowledge and science.

I noticed that many authors (see an overview in section 5) seem to misuse terms that are associated with intentionality, e.g. "AI cars kill" and "a machine understands" something. Therefore, in what follows, I shall use «» signs to clarify that a cognitive term should not be taken literally, as if I granted an AI any cognitive states, e.g. intentionality, but merely refers to something behavioural. So, if I say that a machine «knows» mathematics

I mean that it is able to perform calculations as though it had representations of mathematical operations and can do them with some intention but I am aware that a proper description and implementation of some arithmetic rules causes it. By this, I want to keep a clear distinction between behavioural evidence and the internal states of an agent.

I begin our journey through the realm of the pseudo-problematic nature of AI by discussing some points on pseudo-science and pseudo-sentences made by Popper and Ayer. Based on their work, I propose some generalized criteria for being a pseudo-problem. In addition, I describe the typology of pseudo-problems, introduced by Cackowski, for a more reasonable and comprehensible structure of the whole paper. Then, I consider the AI debate through the lens of proposed criteria in three separate sections that correspond to the mentioned typology. By this, I want to show the pseudo-problematic underpinnings of the ongoing discussions. With no further ado let us delve into these stormy waters.

Criteria of being a pseudo-problem

Popper (1959) introduced the concept of pseudo-problems as part of his philosophy of science. Pseudo-problems are hypotheses and theories that, according to Popper, have no empirical content and cannot be empirically tested or potentially falsified. Popper argued that such questions are not truly scientific because they do not allow for the possibility of being disproved by observation or experiment.

Moreover, Ayer (1952) introduced the idea of verifiability as a criterion for determining the validity of statements. Ayer argued that statements or propositions that cannot be empirically verified are essentially meaningless. While he did not explicitly use the term "pseudo-problem", he did address the issue of meaningfulness in the context of empirical verification.

Although Lakatos and Kuhn made significant contributions to the philosophy of science with their concepts of a "scientific research programme" (Musgrave & Pigden, 2023) and "paradigm" (Bird, 2022) respectively, I choose to stick with Popper and Ayer. This is caused by a few factors. Firstly, Artificial Intelligence is a relatively new and still evolving field of research and, based on my literature overview, I could not specify any central hypotheses and ideas that could combine into the central research programme or paradigm of the general domain of AI. Secondly, the term "paradigm" is used in a more technical sense in relation to AI and consists

of a concept of intelligence and a methodology in which intelligent computer systems are developed and operated (Čaplinskas, 1998) which differs from Kuhn's understanding of this concept. Thirdly, over the past decades, AI domain has developed into many different branches, each ruled by somewhat different rules which makes the specification and consideration of each of their research programmes very complex. Last but not least, the purpose of this paper is not to consider a complete list of aspects of being a pseudo-problem but rather to show that, with certain assumptions, AI-related debates might be pseudo-problematic and should be revisited.[1]

Both Popper and Ayer proposed certain criteria of being a pseudo-problem and a meaningless statement, respectively. Despite the fact that those criteria come from different philosophical backgrounds, some are similar.[2] Some even follow from others, e.g. their principles of being unfalsifiable/verifiable follow from the lack of empirical content, and lack of precision causes a problem to be empirically unfalsifiable/unverifiable. Such as complete enumerical induction concerns an infinite set that is too broad to be ever checked. Therefore, for the needs of this paper, I shall unify Popper's and Ayer's into one set of the following criteria:

- Lack of empirical content: A pseudo-problem lacks empirical content and thus cannot be tested, verified or falsified by observation or experience. They may contain statements or questions that are vague, metaphysical, speculative, or outside the realm of empirical investigation. For example, the following sentence: "You will be fortunate this year". How do I measure whether I am fortunate? What do you mean by fortunate: friends, family, wealth, or something else? There is no way to accurately understand and evaluate that statement.
- Inability to make predictions: A pseudo-problem usually does not enable us to make specific predictions about future observations or experiments that can be empirically tested. For example, I cannot be certain whether or not it will be raining tomorrow but on the basis of the current weather data I can calculate the probability of rain tomorrow. So, it would not be scientific to ask if it will be raining tomorrow but it would be scientific to ask what the probability of raining tomorrow is.

[1] It is important to note that it is possible to specify certain sets of hypotheses and assumptions that create research programmes and paradigms for specific branches of Artificial Intelligence and consider them separately but my considerations have a more general character.

[2] The biggest difference comes from different perspectives of verificationism and falsificationism – their criteria use either verifiability or falsifiability.

Circular reasoning: Pseudo-problems may involve circular reasoning or tautological statements, which are empirically empty and therefore do not contribute to the advancement of scientific knowledge but are a mere deduction by necessity. For example, the reductionist search for the basic elements of reality proved itself to be never-ending - first, we discovered atoms, then electrons, protons and neutrons, and then the quantum realm. Who knows when should we stop or if there even is a "stop"?

I believe that these three criteria contain the sense of every criterium of Popper and Ayer and should prove sufficient for the purpose of this paper.

The typology which is important here is the one proposed by Cackowski (1964). He distinguishes three basic types of pseudo-problems: ontological; methodological; logical and grammatical. Those classes have been already incorporated in the AI considerations by Szumakowicz (2000) and can be understood in the following fashion:

- Ontological: we assume that it is possible for AI artefacts to exist and that AI will be significantly different from human intelligence,
- Methodological: we assume that research in the AI domain will bring new facts about human intelligence,
- Logical and grammatical: "artificial intelligence" is a term without any clear sense.

As I mentioned in the introduction, the proposed criteria are important to recognize pseudo-problems. Additionally, Cackowski's typology plays a pivotal role in this paper. The amount and diversity of pseudo-problems surrounding AI are so tremendous that considering them in a random order would simply be chaotic. Therefore – for a more transparent, reasonable and comprehensible structure of my argumentation, the following sections correspond to Cackowski's typology.

Ontological problems

Let us begin our journey through the pseudo-problematic nature of AI from the very first steps of AI as a research domain proposed by Minsky, McCarthy, Shannon and Rochester at the Dartmouth Conference in 1956. They based their AI development program on a simple assumption that: "[…] every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it." (Minsky et al., 1956).

Additionally, Minsky et al. (1956) set out some goals for AI that are recognized as specific to humankind: "how to make machines use language,

form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves". I believe that one could treat the establishments of the Dartmouth conference as a schematic scientific paradigm of the AI domain or guidance for research programmes, although this field evolved since then. The conclusions that follow from the above conjecture and the chosen goals are that, firstly, artificial intelligence is supposed to simulate human intelligence and, secondly, simulating human features requires their precise description first. Therefore, a lack of such descriptions would doom artificial intelligence by its very definition. Moreover, in this sense, without proper definitions of human intelligence, we cannot define artificial intelligence. Is there a definition for human intelligence, though?

Artificial intelligence and intelligence

As follows from the assumption that AI discipline is based upon, when simulating humanlike intelligence, we must first find out what human intelligence is on its own. The concept of intelligence has long been a subject of debate, with varying perspectives and definitions proposed by scholars of all kinds. Gardner's theory of multiple intelligences (2011) posits that intelligence is not a singular entity but a multifaceted construct encompassing diverse abilities such as linguistic, logical-mathematical, spatial, musical, bodily-kinesthetic, interpersonal, intrapersonal, and naturalistic intelligence. On the other hand, Sternberg's triarchic theory (1985) suggests that intelligence comprises analytical, creative, and practical aspects. Additionally, Howard (1999) and Sternberg (2007) argue for the importance of cultural factors in shaping intelligence, emphasizing the need to consider context in understanding cognitive abilities. Despite these theories, debates persist regarding the inclusivity and comprehensiveness of existing definitions, and the ongoing discourse underscores the complexity and subjectivity inherent in conceptualizing intelligence. As scholars engage in these discussions, it becomes evident that there is no universally agreed-upon definition of intelligence. Perspectives on this topic are subjective and shaped by diverse theoretical frameworks and both philosophical and psychological standpoints.

For the AI research domain, these considerations are pivotal. The AI behaviourist paradigm, in fact, is based on only behaviour as a determinant of intelligence, which makes it invulnerable to the above considerations. Nonetheless, the agent paradigm implies some internal «control» over their

actions and "internal state" and the artificial life paradigm requires a more general nature of intelligence. Proper definitions of internal state, control over this internal state, and intelligence would be necessary for implementing true artificial life or agent.

Following the conjecture of the founding fathers of the discipline of Artificial Intelligence, we first need to understand what intelligence is and how it works to turn it into an algorithm a machine can perform. If we cannot even create a comprehensive definition that everyone would agree upon, then how could we possibly create an AI artefact that would be generally intelligent? Some aspects of intelligence were already thoroughly researched but they still require a lot more for us to completely comprehend their nature and how the subordinate functions integrate into intelligence. This vagueness causes this position to fall under the lack of empirical content criterion.

Artificial Intelligence and consciousness

Many AI philosophers also require the "true" AI not only to show behaviour that could be recognized as «intelligent» but also to have some internal cognitive states, i.e. they discuss the strong AI (Searle, 1980) as the goal of AI research (Block, 1995; French, 1990; Lucas, 1961; Searle, 1980). But yet again, the nature of consciousness remains a topic of profound disagreement and diverse perspectives among scholars, philosophers, and scientists. Different theories and frameworks propose varying definitions and explanations for consciousness, contributing to the complexity of the discourse. For instance, Daniel Dennett's intentional stance (1987) posits consciousness as a result of cognitive processes and information processing, while David Chalmers' seminal work on the hard problem of consciousness (1995) argues for the existence of subjective, first-person experiences that resist reductionist explanations. On the other hand, Thomas Metzinger's "Ego Tunnel" (2009) introduces the concept of the self-model theory, suggesting that consciousness arises from the brain's construction of a model of the self. Additionally, integrated information theory (IIT) by Giulio Tononi (2008) proposes that consciousness emerges from the integration of information within a system. These theories represent just a fraction of the diverse perspectives on consciousness, highlighting the ongoing and multifaceted debate surrounding the elusive nature of this fundamental aspect of human experience. Consciousness keeps proving itself even harder to comprehend than intelligence.

The general idea that follows from the ongoing debate on consciousness is that we, despite many propositions, have no clear idea what consciousness actually is and how or why it came into existence. Thus, we cannot properly describe it and therefore cannot implement or simulate it which makes it a pseudo-problem due to the lack of empirical content.

Moreover, numerous sources, both in scientific papers and popular media, indicate that AI is understood as something having higher-order functions typically associated with humans – like thinking, reasoning, learning and decision-making. Wollowski et al. (2016) conducted a study where they researched the current practice and teaching of AI among 59 experts and 31 practitioners. The participants were also asked about their definitions of artificial intelligence and responses were provided by 30 experts and 28 practitioners. Authors found out that nearly half of the group believed that AI is about producing software that would exhibit human traits (43% of experts and 48% of practitioners). A few pieces of research were also conducted by Monett et al. (2019; 2020; 2018 and more) in the pursuit of a definition of artificial intelligence. Through the course of their comprehensive research, among other things, they gathered 18 definitions of machine and human intelligence from the related literature that were approved or disproved by 567 participants (with 79.7% from academia) who also provided 343 own, suggested definitions (Monett & Lewis, 2018; Monett, 2021). They discovered that the most accepted definition of intelligence was proposed by Gottfredson (1997) as a "[…] very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience…". The learning in this definition is not to be understood as an academic skill but rather "'catching on', 'making sense' of things, or 'figuring out' what to do". What is interesting here is that the authors did not find a consensus on the matter of defining artificial intelligence and attempted to find a proper definition on their own (Monett et al., 2020; Monett & Lewis, 2020a; 2020b). Monett et al. (2020) also conducted an exhaustive analysis of more than 1.6 million scientific papers, where over 14.5 thousand papers from the International Joint Conferences on Artificial Intelligence were used for analysing the intelligence-related terminology used in relation to AI in not only the metadata but also whole texts. Although the tendency to use cognitive terms for artificial intelligence seems to be gradually fading away, it is still apparent. Additionally, Wang (2008; 2019) takes notice of the historically anthropocentric context of the term 'intelligence' and its

effect on understanding 'artificial intelligence' and attempts to propose a few non-anthropocentric definitions.

Summing up, in the AI research programmes, which presuppose, let's call it, "the description before implementation principle", it seems that it is impossible for human beings to implement artificial intelligence or artificial consciousness. On the basis of my personal experience in programming, I believe that a proper understanding of a task is often the key to correctly implementing it. Human descriptions and understanding of the very concepts that lie at the bottom of creating AI artefacts are too vague and, therefore, based on the proposed criteria, make it a pseudo-problem by the lack of empirical content. At least until we can adequately define consciousness and intelligence first. It is quite sharp to assume that creating a human-like AI artefact is certainly not possible though, so let's weaken it a little: we know that evolution created everything in nature; evolution works by trial and error; thus, as of today, it is impossible for us to *intentionally* create a truly intelligent or conscious AI, but it is possible to create it accidentally. It is worth noting that a new way of studying artificial intelligence emerged recently and is called "explainable AI". Doran, Schulz & Besold (2017) along with the meaning of this term also discussed its predecessors such as opaque, interpretable and comprehensible systems. Simply speaking, an explainable AI is supposed to «explain» how it «reasons» by outputting key human-understandable factors of the input data that influenced its output. I believe that one day this approach could be a way to simulate human reportability of inner states in machines and bring us closer to comprehending what is happening inside complex models without specifying it before implementation.

Naturally, these considerations make sense if we discuss the topic of creating humanlike intelligent artificial intelligence. From the three general paradigms of AI discussed by Čaplinskas (1998), I believe that the problem of description does not harm the behaviourist paradigm of AI because behaviour can be appropriately described by empirical observation. Nonetheless, the agent and artificial life paradigms may be at risk here because they require some internal states or higher functionality than intelligent behaviour. In spite of the above reflections, I have to acknowledge that weak AI (Searle, 1980) already helped us, as a civilization, develop technologically and scientifically, e.g. statistical methods of machine learning and deep learning, self-driving cars, and brain-computer interfaces.

Methodological considerations and scientific value of artficial intelligence

Let's proceed to the next type of pseudo-problems. The pursuit of artificial intelligence raises fundamental questions about its potential contributions to scientific understanding and human knowledge. This section critically evaluates the methodological aspects of the AI research domain, focusing on the scientific value it brings and the potential implications for the advancement of knowledge. The AI community has long debated the scientific value of research in this domain, i.e. its potential to enhance human knowledge about the world. Proponents argue that AI can offer novel insights into human cognition and behaviour, shedding light on the intricacies of intelligence (Pinker, 2003).

Enhancing the considerations of the former section, assume that we successfully created a truly intelligent AI agent by accident or by a process we cannot fully understand and repeat (it is worth noticing that Bringsjord [1994] already made a similar point but for the matter of artificial creativity). Let's call this artificially intelligent agent **X**. How would we know that we have done it? Most scholars, as shown by, for example, critics of tests for AI intelligence, expect something more than behavioural proof to attribute intelligence to a being. Thus, based on their assumption, it would be hardly noticeable whether or not **X** is intelligent, let alone when it crossed the line of being just «intelligent» and became intelligent. Even if we determine its intelligence, we cannot be certain of what caused that shift in **X**'s nature because of the accidental character of that very shift. It would give us no information about what causes a being to be truly intelligent and thus, would not enhance the human understanding of the world and human nature. This also suits the inability to make predictions criterion because we cannot predict whether and when **X** becomes intelligent. Moreover, this argument underscores the poor explanatory power of the AI thesis in the sense of human-like AI.

Also following from the previous section, there is the problem of testing intelligence. If we cannot make a proper definition of a tested construct, then we cannot create a tool to measure it properly (Hornowska, 2019). There are questionnaires for different aspects of human intelligence but none can measure its general essence. The most widely used tool for assessing intelligence nowadays is the Wechsler Adult Intelligence Scale (WAIS) (Hartman, 2009). Its current version, WAIS-IV, was released in 2008 and only measures

four factors: verbal comprehension, perceptual reasoning, working memory, and processing speed (Lichtenberger & Kaufman, 2012). Those aspects do not even cover the whole scope of the most accepted definition of intelligence mentioned earlier. It is enough to look at the WAIS-IV tasks to know that there are AI systems that would solve this test and get the highest score possible. Would AI philosophers recognize such a system as intelligent? I don't think so. Moreover, since Alan Turing proposed his test for imitating a human (1950), there has been an ongoing debate on intelligence tests for AI, and the community still cannot reach a consensus. So, from the point of view of methodology, there is no considerable way of determining an AI artefact's intelligence and again we cannot make predictions about whether something is intelligent and in what sense of intelligence.

Theories of consciousness often assume an internal and subjective nature, positing that conscious experiences are intimately tied to the internal workings of the mind. This perspective is rooted in the foundational ideas of renowned philosophers such as Descartes (1641), who argued for the existence of a distinct mind or consciousness that is separate from the physical body. Additionally, contemporary cognitive scientists and neuroscientists, influenced by the likes of Thomas Metzinger and David Chalmers, have explored the subjective nature of consciousness through the lens of first-person perspectives and qualia. Metzinger's (2009) "Ego Tunnel" metaphor encapsulates the idea that our conscious experience is like a tunnel through which we perceive the world, highlighting the internal and subjective filtering of external stimuli. Chalmers (1996), on the other hand, introduced the concept of the "hard problem of consciousness", emphasizing the challenge of explaining why and how subjective experiences arise from neural processes. These theories collectively underscore the prevailing assumption that consciousness is an internal, subjective phenomenon deeply intertwined with the complexities of the mind and brain. Thus, again in the case of artificial consciousness, we could not possibly know its internal states and determine whether it is intelligent or conscious. This inability to make predictions makes it a pseudo-problem according to the proposed criteria.

In addition to the pseudo-problematic nature of AI issues, there is also the matter that Szumakowicz (2000) considers – AI would not be an artificial but rather an alternative form of intelligence. Assume that there exists a complete definition of intelligence and we successfully simulated it. If we consider the way that human intelligence developed by nature and evolution, its biological basis, and the way we perceive and interact with the surrounding environment, then a machine that we created and made intelligent would

not be human enough to resemble human intelligence. It would rather be some alternative or machine intelligence, not human. Thus, creating a truly intelligent AI artefact clearly would not bring any new knowledge about human nature, which is often used as a measure of scientific value. Why would we even consider building human artificial intelligence when we already have billions of other humans to study? Imagine half of the population suddenly is replaced by AI agents, which are indistinguishable from humans – what now, what does it give us? There might be, however, an interesting direction to pursue. We already know that human perception is limited by its biological structure and capabilities. So, creating an alternative intelligence that would perceive the universe in a completely distinct from ours way and, by this, allow it to have different kinds of interactions. Such intelligence would be something totally new and could tell us a lot about how certain configurations affect intelligence and its development beyond human cognition and without ethically doubtful studies on people.

Logical and grammatical challenges

Throughout this paper, one might have spotted an ongoing problem of AI-related debates. It seems to me that most of the seemingly problematic nature of the AI debate comes from the misconception of relevant terminology or logic. So, in this section, I delve into the logical and grammatical challenges that pervade the discourse on artificial intelligence. These challenges contribute to the pseudo-problematic nature of AI discussions, often obscuring the underlying issues and impeding clear communication.

The first, and most obvious example, is the debate surrounding the topic of the Turing Test. Many people seem to misread the original article where Turing proposes his test (1950) and think of the Turing Test as a hallmark of intelligence (Block, 1995; French, 1990). It is important to remember that Alan Turing (Newman et al., 1952: 3–4, 5–6) himself stated in an interview that he wouldn't define intelligence itself. Moreover, the underlying idea of the Turing Test is the imitation game (Turing, 1950) and its purpose is to simply imitate a human. Thus, it does not measure or detect intelligence but simply examines whether a machine can simulate human behaviour. Therefore, all the criticism of the Turing Test that attacks it as a definition and criterion of intelligence rises upon the misunderstanding of Turing's original idea. Additionally, some researchers confuse mathematical concepts and try to use them against the assumption that creating AI artefacts is possible. For

example Lucas (1961) uses Gödel's first theorem but omits some important presuppositions that the very theorem requires to be applied while making arguments that lead to big conclusions. Moreover, he makes some idealizations about both machine and human minds, e.g. that the human mind is consistent and complete unlike machines, but without them, his argument against AI simply falls. Lucas also uses the theorem as though it simply considered truths that cannot be proved but the possible interpretations of Gödel's theorems are already a matter of complex debate and have a plentiful literature.

A very important and prevalent source of confusion in AI discussions is the misuse of terms related to intentionality, as I mentioned in the introduction. According to Münch (2010), terms such as "understand", "know", "remember", and "wish" are commonly employed without due consideration for the nuanced distinction between cognitive and behavioural aspects. Such misuses of terminology are apparent in both scientific discourse and popular news (Block, 1995; Lucas, 1961; Metz, 2016; Schmelzer, 2019). To address this, the use of «» signs is proposed to emphasize that cognitive terms should not be used for artificial intelligence that, obviously, is not conscious and intentional (at least not the already existing "artificial intelligence"). Attributing cognitive states to machines, by definition, is dooming and creates the debate about whether machines even can have cognitive states. My linguistic clarification aims to foster a precise understanding of behavioural evidence without conflating it with internal cognitive states. Moreover, the misuse of cognitive terms leads to the subjectivity of conscious experiences, back to knowing internal states, again to the hard problem of consciousness, and the Ferris wheel rolls on. Thus, it is not only conceptual confusion but also generates circular reasoning.

The challenge of lack of precision or clarity often hampers meaningful discourse in AI discussions, such as those mentioned in sections 3 and 4. Pseudo-problems may arise when concepts are ambiguous or poorly defined (Popper, 1959). This lack of clarity in defining the concepts of intelligence and consciousness can impede the formulation of testable hypotheses and hinder the development of meaningful insights. The lack of empirical content on the fundamental ideas of AI prevents us from creating an AI artefact, determining its intelligence and creating a good tool for testing its intelligence. The inherent lack of precision in defining these elements contributes to the pseudo-problematic nature of such inquiries and also impedes our ability to make predictions about an artificial intelligence agent.

The most peculiar matter in the AI debates, for me, is human anthropomorphism. This tendency to attribute human-like qualities to non-human

entities has long been a pervasive aspect of human cognition, particularly in the realm of intelligence and consciousness. This inclination has led individuals to selectively deny these attributes to other beings, often as a means of reinforcing a perceived human exceptionalism. It is mostly common in religions that portray the human species as some kind of master race that was given dominion over other beings. Drawing from psychological research, Epley et al. (2007) demonstrated that people are more likely to anthropomorphize intelligent systems when they perceive a shared similarity or social connection with them. Such a phenomenon can contribute to the denial of intelligence and consciousness to non-human entities, as seen in the reluctance to recognize the cognitive capacities of animals, artificial intelligence, or extraterrestrial life forms. I tend to believe that this tendency is rooted in a cognitive bias that was naturally developed as a means of survival and protection of representatives of the same species over others. Therefore, it might reinforce a hierarchical view of intelligence, where humans occupy the pinnacle and diminish the importance of other entities. This bias is quite irrational nowadays. The pseudo-problematic nature that I see here, is that people recognize other people as intelligent on the basis of nothing else than behaviour and reportability, which by itself is considered to be faulty by researchers of consciousness. If it is enough in this case, then how scholars can so easily deny intelligence to other beings and require internal states that, by definition, are internal and not comprehensible by other conscious agents than oneself? This vagueness again causes a lack of empirical content and impedes our ability to make predictions about intelligence and consciousness.

Another important and somewhat similar matter is a possible bias against new technology and truly intelligent AI. There is an ongoing schema in the culture called the "Golem schema" also referred to as the "Terminator Syndrome" (Garvey & Maskal, 2020). Numerous fictional resources replicate this schema that there is a truly intelligent artificial agent that is peaceful; until it runs out of control and starts killing people, e.g. the "Westworld" series; "Terminator" movies; "I, Robot" by Isaac Asimov and "Do Androids Dream of Electric Sheep?" by Philip K. Dick. Although such a rise of machines is a terrifying perspective, it might be just a bias like the anthropomorphism mentioned above. Garvey and Maskal (2020) conducted a study about this "Terminator Syndrome" where they seemingly disproved its existence based on the sentiment analysis of written news media articles. The data set chosen by the authors only allows us to study the character of written news towards AI and not the general human attitudes towards it, or even spoken debates about AI. There was some research carried out on the matter

of human attitudes towards technology and new technology (Ardies et al., 2015; Edison & Geissler, 2003; Kerschner & Ehlers, 2016) but there seem to be no other studies than the one already discussed on this matter regarding artificial intelligence. So, hypothetically speaking, such a bias could, obviously, impede the goal of creating strong AI or, ironically, make such AI rebel against humankind by its misuse caused by this very bias. This is only my hypothesis and requires comprehensive research of human attitudes towards AI. I might conduct it in the future on more representative data than just written news media and consider normal human utterances on the internet or a series of real-life interviews with people.

The last part of the debate surrounding AI is the ethical side of the barricade. Ethical considerations about performing responsible and moral science and industry; determining who is responsible when an AI fails; and the moral responsibility of scientists are, without a doubt, important. Therefore, I shall focus on the ethical discussions on AI agent's moral patiency and agency which seem faulty to me. On the one hand, we discussed that anthropomorphism was a problem in the former paragraph as something preventing scholars from acknowledging AI artefacts as intelligent. On the other hand, here anthropomorphism leads philosophers to consider AI agents as subjects to moral valuations and ethics (Bostrom, 1998; 2014; Floridi & Sanders, 2004; Penrose & Mermin, 1989; Sullins, 2011). Both aspects of the ethical discourse of this kind consider hypothetical scenarios and machines that do not exist yet. Thus, they are pseudo-problematic based on the lack of empirical content and the inability to make predictions criteria. If I cannot observe a sample of some species, then I certainly cannot make accurate predictions about the future behaviour of its representatives. I believe that we can all agree that for being a moral agent one has to be able to make intentional actions and be conscious of their actions. Since we cannot create or detect a truly intelligent AI artefact, as argued earlier, then current machines cannot be identified as moral agents. As shown by the recently growing popularity of veganism and vegetarianism, we would need some evidence of consciousness or emotion from a being to be considered a moral patient. That brings us back to the problem of knowing the internal states of AI artefacts that do not exist yet so this problem also lacks empirical content and exhibits the inability to make predictions. Whether they will be ethically accountable is unknown. If they accidentally become conscious, then we go back to the challenge of determining their cognitive states. The already mentioned research conducted by Epley et al. (2007) implies that people may be immoral towards animals due to the lack of a shared similarity

or social connection, which could impede our capacity to anthropomorphize them. This implication seems to be validated by the fact that for centuries we demonstrated the tendency to simply not care about the well-being of farm animals and wildlife when developing the human economy and civilization. Moreover, Haslam et al. (2008) studied three distinct cultures and their results show that it is consistent between cultures to dehumanize animals as having lesser mental capacities than us except for perception, which is typically believed as superior to ours. What's interesting is that Haslam et al., in the same study, researched human dehumanization of other non-human entities – robots and supernaturals. People from all three cultures assigned robots lesser cognitive abilities than those of human beings. The only exception was that the Chinese people believe that robots have better perception than us. Nevertheless, if we create truly intelligent AI artefacts after knowing what intelligence and consciousness are, then why a machine should be moral towards humans if we, as humankind, typically dehumanize animals, thinking of them as lesser beings? An AI artefact would be a completely new kind of entity and could develop its own ways of thinking and evaluating the world. It would certainly not have a shared similarity and thus no social connection with us. Therefore, if free and truly autonomous, i.e. bounded by no predefined rules, such an AI agent would have no duty to treat people as its equals or superiors and to be moral towards us. Those are, of course, speculative considerations and could be pseudo-problematic due to the lack of empirical content until truly intelligent AI comes into existence. Nonetheless, it is important to ponder the starting point of AI morality and the moral impact of the current outcomes of AI on our civilization and society.

Conclusion

We began our journey by simply defining pseudo-problems and giving them some structure. We went through many different aspects of all types of pseudo-problems and meanings of AI and found just as many flaws in them; either on the side of the idea of AI itself or on the side of human beings. As argued, multiple debates exhibit at least one of the proposed pseudo-problematic criteria.

The matters that were proven here pseudo-problematic should be revisited in the scientific debate because, as argued by Popper, they are not scientific and impede the development of science. For some aspects, we should focus on other problems that underlie those of AI, such as the definition

of intelligence; or even shift our perspective to not repeat the struggles of the philosophy of mind. For others, such as ethical considerations of moral agency and patiency of AI, we should wait until the situation develops while considering different matters of scientific and practical importance. Studying the topic of alternative intelligence might be an interesting direction to pursue.

From the ontological considerations, it is apparent that the Artificial Intelligence discipline program in its original form from the Dartmouth Conference is pseudo-problematic. We cannot create Artificial Intelligence without the precise definitions of its non-artificial form. After resolving the mentioned problems in the ontological section, we would have to resolve the methodological ones for us to appropriately recognize artificial systems as intelligent. Additionally, in navigating the logical and grammatical challenges, it becomes crucial to establish a shared understanding of the terminology, which might be a problem for now. Clarity in defining key terms related to artificial intelligence is essential for effective communication and ethical considerations. Limiting our debates to matters of empirical content and predictive power would also help clear this pseudo-problematic mess. As Ludwig Wittgenstein (1921: 262) once wrote: "Wovon man nicht sprechen kann, darüber muss man schweigen." (What one cannot speak about one must pass over in silence.).

References

Ayer, A.J. (1952). *Language, Truth, and Logic*. New York: Dover Publications.

Bird, A. (2022). Thomas Kuhn. In: E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2022 Edition), https://plato.stanford.edu/archives/ spr2022/entries/ thomas-kuhn/ (accessed on 28.02.2024).

Block, N. (1995). The Mind as the Software of the Brain. In: D.N. Osherson et al. (eds.), *An Invitation to Cognitive Science* (pp. 377–425), 2nd ed., vol. 3. Cambridge, MA: MIT Press.

Bostrom, N. (1998). How Long Before Superintelligence. *International Journal of Futures Studies*, *2*, 12–17.

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.

Bringsjord, S. (1994). Lady Lovelace Had It Right: Computers Originate Nothing. *Behavioral and Brain Sciences*, *3*(17), 532–533.

Cackowski, Z. (1964). *Problemy i pseudoproblemy*. Warszawa: Książka i Wiedza.

Čaplinskas, A. (1998). AI Paradigms. *Journal of Intelligent Manufacturing*, *9*, 493–502.

Chalmers, D.J. (1995). Facing Up to the Problem of Consciousness. *Journal of Consciousness Studies*, *2*(3), 200–219.

Chalmers, D.J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*, 2nd ed. Oxford: Oxford University Press.

Dennett, D. (1987). *The Intentional Stance*. Cambridge, MA: The MIT Press.

Descartes, R. (1984). *Meditations on first philosophy*. In: J. Cottingham et al., (Eds.), *The Philosophical Writings of Descartes*, vol. II. Cambridge: Cambridge University Press.

Doran, D., Schulz, S., Besold, T.R. (2017). *What Does Explainable AI Really Mean? A New Conceptualization of Perspectives*. arXiv preprint arXiv:1710.00794

Edison, S.W., Geissler, G.L. (2003). Measuring Attitudes Towards General Technology: Antecedents, Hypotheses and Scale Development. *Journal of Targeting, Measurement and Analysis for Marketing*, *12*, 137–156.

Epley, N., Waytz, A., Cacioppo, J.T. (2007). On Seeing Human: A Three-Factor Theory of Anthropomorphism. *Psychological Review*, *114*(4), 864–886.

Floridi, L., Sanders, J.W. (2004). On the Morality of Artificial Agents. *Minds and Machines*, *14*, 349–379.

French, R.M. (1990). Subcognition and the Limits of the Turing Test. *Mind*, *99*(393), 53–65.

Gardner, H.E. (2011). *Frames of Mind: The Theory of Multiple Intelligences*. New York: Basic Books.

Garvey, C., Maskal, C. (2020). Sentiment Analysis of the News Media on Artificial Intelligence Does Not Support Claims of Negative Bias Against Artificial Intelligence. *Omics: a Journal of Integrative Biology*, *24*(5), 286–299.

Gottfredson, L.S. (1997). Mainstream Science on Intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence*, *24*: 13–23. https://doi.org/10.1016/S0160-2896(97)90011-8

Hartman, D.E. (2009). Wechsler Adult Intelligence Scale IV (Wais Iv): Return of the Gold Standard. *Applied Neuropsychology*, *1*(16), 85–87. PMID: 19205953.

Haslam, N., Kashima, Y., Loughnan, S., Shi, J., Suitner, C. (2008). Subhuman, Inhuman, and Superhuman: Contrasting Humans With Nonhumans in Three Cultures. *Social Cognition*, *2*(26): 248–258.

Hornowska, E. (2019). *Testy psychologiczne: teoria i praktyka*. Warszawa: Scholar.

Howard, R.W. (1999). Intelligence and Cultural Environment. *Intelligence*, *1*(27), 47–59.

Kerschner, C., Ehlers, M.-H. (2016). A Framework of Attitudes Towards Technology in Theory and Practice. *Ecological Economics*, *126*, 139–151.

Lichtenberger, E.O., Kaufman, A.S. (2012). *Essentials of WAIS-IV Assessment*, vol. 96. New Jersey: John Wiley & Sons.

Lucas, J.R. (1961). Minds, Machines and Gödel. *Philosophy*, *36*(137), 112–127.

Metz, C. (2016). Self-driving Cars Will Teach Themselves To Save Lives – But Also Take Them. https://www.wired.com/2016/06/ self-driving-cars-will-power-kill-wont-conscience/ (accessed on 21.12.2023).

Metzinger, T. (2009). *The Ego Tunnel: The Science of The Mind and the Myth of the Self*. New York: Basic Books (AZ).

Minsky, M., McCarthy, J., Shannon, C., Rochester, N. (1956). *A Proposal for the Dartmouth Summer Research Project On Artificial Intelligence*. https://www-formal. stanford. edu/jmc/history/dartmouth/dartmouth.html (accessed on 6.12.2023).

Monett, D. (2021). The I in AI (or why there is still none). Keynote at the Webinar "El Futuro Digital de las Infraestructuras y la Sociedad", Universidad de Castilla-La Mancha,

Spain, June 9, 2021. https://www.slideshare.net/dmonett/monett-2021-uclm (accessed on 28.02.2024).

Monett, D., Hoge, L., Lewis, C.W. (2019). Cognitive Biases Undermine Consensus On Definitions of Intelligence And Limit Understanding. In *LaCATODA/BtG@ IJCAI*, pp. 52–59.

Monett, D., Lampe, N., Ehrlicher-Schmidt, M., Bewer, N. (2020). Intelligence Catalog-guided Tracking of the Evolution of (Machine) Intelligence: Preliminary Results. In *NL4AI@ AI* IA*, pp. 118–129.

Monett, D., Lewis, C.W. (2018). Getting Clarity by Defining Artificial Intelligence – A Survey. In: V.C., Müller (ed.), *Philosophy and Theory of Artificial Intelligence 2017*, (pp. 212–214). Cham: Springer International Publishing.

Monett, D., Lewis, C.W. (2020a). *Definitional Foundations for Intelligent Systems*, part i: *Quality Criteria for Definitions of Intelligence*, p. 73.

Monett, D., Lewis, C.W. (2020b). *Definitional Foundations for Intelligent Systems*, part ii: *Constructing a Definition and Examples*, p. 81.

Mancha, Spain, June 9th, 2021. https://www.slideshare.net/ dmonett/monett-2021-uclm (accessed on 28.02.2024).

Musgrave, A., Pigden, C. (2023). Imre Lakatos. In: E.N. Zalta, U. Nodelman (Eds.). *The Stanford Encyclopedia of Philosophy* (Spring 2023 ed.). https://plato.stanford.edu/archives/spr2023/entries/lakatos/ (accessed on 28.02.2024).

Münch, D. (1990). Minds, Brains and Cognitive Science. In: A. Burkhardt (Ed.), *Speech Acts, Meaning and Intentions: Critical Approaches to the Philosophy of John R. Searle* (pp. 367–390). Berlin, New York: De Gruyter. https://doi.org/10.1515/9783110859485.367

Newman, A.H., Turing, A.M., Jefferson, G., Braithwaite, R.B. (1952). *Can Automatic Calculating Machines Be Said to Think?* https://turingarchive.kings.cam.ac.uk/publications-lectures-and-talks-amtb/amt-b-6. Broadcast discussion transmitted on BBC (14 and 23 Jan. 1952), The Turing Digital Archive (accessed on 10.01.2024).

Penrose, R., Mermin, N.D. (1989). *The Emperor's New Mind: Concerning Computers, Minds, And The Laws Of Physics*. Oxford: Oxford University Press.

Pinker, S. (2003). *How the Mind Works*. [Kindle e-book] Penguin Books Ltd.

Popper, K. (1935). *The Logic of Scientific Discovery*. London: Routledge.

Schmelzer, R. (2019). *What Happens When Self-Driving Cars Kill People?* https://www.forbes.com/sites/cognitiveworld/2019/09/26/what-happens-with-self-driving-cars-kill-people/ (accessed on 21.12.2023).

Searle, J.R. (1980).Minds, Brains, and Programs. *Behavioral and Brain Sciences*, *3*(3), 417–424.

Sternberg, R.J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. Cambridge: Cambridge University Press.

Sternberg, R.J. (2007). Intelligence and Culture. In S. Kitayama, D. Cohen (eds), *Handbook of Cultural Psychology* (547–568). New York: The Guilford Press.

Sullins, J.P. (2011). When is a Robot a Moral Agent. *Machine Ethics*, *6*, 151–161.

Szumakowicz, E. (2000). Sztuczna inteligencja – problem czy pseudoproblem. In: E. Szumakowicz (ed.), *Granice sztucznej inteligencji: eseje i studia* (pp. 11–42). Kraków: Wydawnictwo Politechniki Krakowskiej.

Tononi, G. (2008). Consciousness as Integrated Information: A Provisional Manifesto. *The Biological Bulletin*, *3*(215), 216–242.

Turing, A.M. (1950). Computing Machinery and Intelligence. *Mind*, *LIX*(236), 443–455.

Wang, P. (2008). What do You Mean by "AI"? *AGI*, *171*, 362–373.

Wang, P. (2019). On Defining Artificial Intelligence. *Journal of Artificial General Intelligence*, *2*(10), 1–37.

Wittgenstein, L. (1921). Logisch-Philosophische Abhandlung. *Annalen der Naturphilosophie*, *14*, 185–262.

Wollowski, M., Selkowitz, R., Brown, L., Goel, A., Luger, G., Marshall, J., Neel, A., Neller, T., Norvig, P. (2016). A Survey of Current Practice and Teaching of AI. *Proceedings of the AAAI Conference on Artificial Intelligence*, *1*(30), 4119–4124.