Proceeding Paper

Ontology and AI Paradigms †

Roman Krzanowski [ORCID] and Pawel Polak *[ORCID]

Faculty of Philosophy, The Pontifical University of John Paul II in Krakow, Kanonicza Street 9,
31-002 Krakow, Poland; rmkrzan@gmail.com
* Correspondence: pawel.polak@upjp2.edu.pl
† Presented at Philosophy and Computing Conference, IS4SI Summit 2021, online, 12–19 September 2021.

Abstract: Ontologies of the real world that are realized, internally, by AI systems and human agents are different. We call this difference an ontological gap. The paper posits that this ontological gap is one of the reasons responsible for the failures of AI, in realizing Artificial General Intelligence (AGI) capacities. Moreover, the authors postulate that the implementation of the biosemiotics perspective and a subjective judgment in synthetic agents would seem to be a necessary precondition for a synthetic system to realize human-like cognition and intelligence. The paper concludes with general remarks on the state of AI technology and its conceptual underpinnings.

Keywords: AI paradigms; GOFAI; embodied AI; biosemiotics; artificial agents; artificial general intelligence; Jacob von Uexküll; ontological gap; intentionality; judgement

1. Introduction

The ultimate goal of AI has long been to construct synthetic minds that can think like humans (e.g., Hibbard, 2002). The three AI paradigms of Good Old-Fashioned AI (GOFAI), Machine Learning (ML) and Artificial Neural Network (ANN) systems, and Situated, Embodied, Dynamical (SED) systems (see also, for example, [1]) have realized only some of the AI's original objectives (e.g., [2]).

AI systems fall short of these original goals due to how they relate to the real world [3]. Ontologies of the real world that are realized, internally, by these three AI paradigms are disconnected from the world by an ontological gap. In other words, the way these systems internally represent the external world does not match, exactly, the way the world is. This is the case as AI systems (using any of these three AI paradigms) do not cognate reality the way humans do, thus, they do not build the internal world ontologies the way humans "build" ontologies of reality (e.g., [4]).

2. Ontologies of AI Paradigms

In GOFAI systems, the world is represented as a collection of clear, well-defined concepts, along the lines of Descartes' "clear and distinct ideas" [3]. Machine learning (ML) and artificial neural networks (ANN) search through a massive number of digitized strings that represent fragments of the world (e.g., [3]), for sub-symbolic probabilistically weak correlations. ML systems can certainly recognize patterns. However, these patterns are 'co-structured' from relations rather than from objects. Despite this limitation, ML constructs have been much more successful than the first wave of AI systems. SED systems, meanwhile, engage with the world through mind, body, and environment (e.g., [5,6]). They partially replicate three aspects of the human cognition of "being in the world"; namely, concrete action, situatedness, and interactionism [7]. Nevertheless, the outside world still does not matter to them as it matters to us, because their grounding in reality is inconsequential to them.

3. Ontologies of Human Agents

Human agents interpret the world differently. The reality presented to us through the senses is not Cartesian, as it is in GOFAI. It is messy, cluttered, "de-ontologized", and meaningless (e.g., [8,9]). Indeed, human consciousness creates an internal ontology (it does not receive it). In other words, it creates a conceptual representation of the external world (e.g., [3]). Our conceptualizations of the external ontology are more or less accurate, because this is a prerequisite for safely engaging with reality. It also means we take responsibility for the ontology we create. We can distinguish between our representation and the external world and acknowledge the difference between them and what this means for us. In other words, we are aware, in principle, of the ontological gap that exists, at least most of the time, but synthetic systems are not. They cannot separate internal representations from external objects, so they do not assume responsibility for their ontology [3].

Let us consider, for example, an autonomous car. To an autonomous car, a pedestrian is a virtual object (represented as a state of some part of memory), an obstacle to be avoided (a programmed constraint), rather than a person who needs to be treated as such. Thus, an autonomous car running into this "person-object" violates programming constraints. But a human driver running into the same "person-object" violates (or should) his/her ethics, morality, violates written and unwritten rules of society, violates laws, and opens the prospects of severe punishment and other severe consequences. Obviously, an autonomous car does not feel the weight of its actions in this way. As previously mentioned, the difference is in ontologies.

4. Ontological Gap between Artificial and Natural Agents—A Biosemiotic Perspective

Any conceptualization of ontology results in an ontological gap between the concepts and the reality being described. In the case of artificial systems, however, we face a very specific difference—the gap between artificial agent and human agent ontologies.

A good explanation of the differences between a human agent's ontology and the ontology of an artificial agent come from the biosemiotic perspective [10]. In this perspective, a subject's bodily/physical engagement with reality enables him/her to identify (and extract) the meaningful parts of reality, the parts of reality that are then built into the agent's internal ontology. These ontologies are the basis for the construction of a subject's internal world of meaning, and biosemiotics theory is an advanced attempt to explain how biological organisms actually construct their ontologies and associated realms of meaning. Thus, artificial agents, which implement the biosemiotic paradigm (for example, through application of Jacob von Uexküll's concept of *Umwelt*, *Innenwelt* and functional circles) could, in some sense, replicate a human agent's engagement with reality and build the human-agent-like internal ontology. It seems that without the application of a biosemiotic approach to artificial intelligence, we have little chance of solving the problem of grounding the symbols of artificial entities and, thus, even SED systems are unable to autonomously produce their own internal ontologies in the way human agents do.

The biosemiotic approach (which assumes embodied, actively and autonomously acting entities in the environment), if true, also delineates important, impassable differences in the ontologies created by artificial agents and by humans. The differences in physical (bodily) constitution, entailing significant differences in behavior, seems to be the factor limiting the similarity of the ontologies of artificial agents and humans.

5. Ontology and Judgements of an Agent

The ability to build its own ontology of the world is certainly a key to creating human-like Artificial General Intelligence (AGI). However, the very ability to produce its own ontology presupposes a specific relationship of the agent to reality. In philosophical language, we call it intentionality. Philosophical reflections on the subject's relation to reality (e.g., Brentano, Twardowski, Husserl, see [11]), have linked intentionality to the concept of judgement, which we try to adopt, here, to artificial agents.

When engaging with reality, the characteristics of the human mind include the property of subjective judgment. Judgment is a technical term here, and it is understood as an agent's capacity to distinguish between the internal and external world, and the capacity to commit to its own existence, wellbeing, and ontology. In this way, the concept of judgment allows us to describe the intentional orientation of the subject towards the external world, and is a key concept of the philosophy of the subject (see e.g., [12]).

We may postulate that the application of the biosemiotic perspective plays a crucial role in the possibility of the creation of autonomous ontologies in artificial agents. It is a base for implementation of an ability to formulate subjective judgment in synthetic agents. The latter is claimed to be a necessary precondition for a synthetic system to cognize like us [3].

6. Conclusions

A coda to this paper may be coming from Wooldridge's book on AI [13] (see also [14]). Wooldridge's message is about the failure of AI to develop the proper philosophical foundations behind the AI quest. However, it equally applies to failure of the proper grasp of ontological gap (as discussed in this paper) between the ontologies of the real world, as realized internally by AI systems and by human agents.

We face, according to Wooldridge, the fact that, thus far, all AI revolutions have failed to achieve the dream of AI (i.e., AGI). This is the yardstick by which we can judge AI and its status and progress, or lack of it. AI systems have so far failed because they have been based on incorrect, oversimplified assumptions about the mind, reality, reasoning, thinking, and human nature and its embodiment in the world. We have developed AI systems that exceed human capacities, but only in specific, narrowly defined tasks. Even if these capacities appear immensely impressive to us, they are not what we are seeking.

We have begun developing AI under Descartes' shadow, with Descartes' views about reality being composed and presented to us in clear, logical, distinct ideas. Turing was also wrong, because people do not only "reckon" as he thought. This was the first and second wave of AI. The reality is not that of Descartes, nor is our thinking like that of Turing's machine. The reality is messy, unpredictable, and fluctuating. Clear distinct ideas and logical reasoning only go so far, if they go far at all. The realities of life are too complex to be programmed in clear, distinct steps (combinatorial explosion), and the ontology of reality is not that of the box world. If we see the world's problems like stacking boxes, to reach some hanging bananas, we only get bananas and little else. Knowledge is not a finite set of prescribed, deterministic, static rules, complex or otherwise. This is why the second AI wave with its knowledge systems failed. With neural nets, we obtained self-learning systems, yet comparing artificial neural nets to the structure of the brain is, again, a mistake. Human neural systems do not need millions of cases to learn how to recognize a cat, for example. We do not actually have a theory of the mind that would be implementable or even one that would explain "how the mind does it". Nor do any reductionist theories add up to the human mind, as they are reductionist theories, at least at present (even given the very interesting attempts to formalize the issue using category theory [15]). Therefore, the failure of AI technology is a failure of our philosophy. We have theories, but we simply still do not know how the mind works. AI research seems a bit like groping in the dark.

However, it also seems that the technical solution to these conceptual problems is possible. We should see the ontological gap and the shortcomings of the philosophical underpinning of AI as technical failures, i.e., failures that can be resolved, given time, once we realize what is missing in our current systems. The proper approach to AI must satisfy two criteria: empirical (i.e., technical) and conceptual (i.e., philosophical) [16]. The empirical criterion requires that the technical solutions and methods must work in real situations. The conceptual criterion requires that the proposed solutions must close the conceptual gap, including the ontological gap, between cognitive sciences and AI technology.

References

1.　Frankish, K.; Ramsey, W.M. Introduction. In *The Cambridge Handbook of Artificial Intelligence*; Frankish, K., Ramsey, W.M., Eds.; Cambridge University Press: Cambridge, UK, 2014; pp. 1–15, ISBN 978-0-521-87142-6.
2.　Russell, S.J. *Human Compatible: Artificial Intelligence and the Problem of Control*; Penguin Books: New York, NY, USA, 2020, ISBN 978-0-525-55863-7.
3.　Smith, B.C. *The Promise of Artificial Intelligence: Reckoning and Judgment*; The MIT Press: Cambridge, MA, USA, 2019, ISBN 978-0-262-04304-5.
4.　Varela, F.J. *The Embodied Mind: Cognitive Science and Human Experience*; The MIT Press: Cambridge, MA, USA; London, UK, 1993, ISBN 978-0-262-72021-2.
5.　Russell, S.J.; Norvig, P. *Artificial Intelligence: A Modern Approach*, 4th ed.; Pearson Series in Artificial Intelligence; Pearson: Harlow, UK, 2020, ISBN 978-1-292-40113-3.
6.　Schöner, G.; Reimann, H. Understanding Embodied Cognition through Dynamical System Thinking. In *The Routledge Companion to Philosophy of Psychology*; Robins, S., Symons, J., Calvo, P., Eds.; Routledge, Taylor & Francis Group: Abingdon, UK; New York, NY, USA, 2019; pp. 505–528, ISBN 978-0-429-24462-9.
7.　Beer, R.D. Dynamical Systems and Embedded Cognition. In *The Cambridge Handbook of Artificial Intelligence*; Frankish, K., Ramsey, W.M., Eds.; Cambridge University Press: Cambridge, UK, 2014; pp. 128–148, ISBN 978-0-521-87142-6.
8.　Zeki, S. *Splendors and Miseries of the Brain: Love, Creativity, and the Quest for Human Happiness*; Wiley-Blackwell: Chichester, UK; Malden, MA, USA, 2009, ISBN 978-1-4051-8558-5.
9.　Bołtuć, P. Conscious AI at the Edge of Chaos. *J. AI Consci.* **2020**, *7*, 25–38. [CrossRef]
10.　Sarosiek, A. The Role of Biosemiosis and Semiotic Scaffolding in the Processes of Developing Intelligent Behaviour. *Zagadnienia Filoz. Nauce* **2021**, *70*, 9–44.
11.　van der Schaar, M. *Brentano, Twardowski and Stout*; Oxford University Press: Oxford, UK, 2016; Volume 1.
12.　Twardowski, K. *Zur Lehre vom Inhalt und Gegenstand der Vorstellungen: Eine Psychologische Untersuchung*; Alfred Hölder, K.U.K. Hof- und Universitäts-Buchhändler: Wien, Austria, 1894.
13.　Wooldridge, M. *The Road to Conscious Machines: The Story of AI*; Penguin: London, UK, 2021, ISBN 978-0-241-33390-7.
14.　Neapolitan, R.E.; Jiang, X. *Artificial Intelligence: With an Introduction to Machine Learning*, 2nd ed.; Chapman and Hall/CRC: Boca Raton, FL, USA, 2018, ISBN 978-1-315-14486-3.
15.　Awodey, S.; Heller, M. The Homunculus Brain and Categorical Logic. *Zagadnienia Filoz. Nauce* **2020**, *69*, 253–280.
16.　Smith, B.C. *On the Origin of Objects*; A Bradford Book—MIT Press: Cambridge, MA, USA, 1996, ISBN 978-0-262-19363-4.