AI AS A POTENTIAL MORAL SUBJECT

When considering the moral subjectivity and the MS of AI, it is necessary to clarify the very concept of AI, which has been a challenging task since its introduction, and some researchers believe that this is an unrealistic goal at the current stage of research. At a general level, there is consensus that AI is the attempt "to make a computer work like a human mind". According to Lindes, the concept "artificial intelligence" should be used in two main senses, which the researcher labels as AI1 and AI2. According to him, AI1 refers to the quality of intelligence in the man-made computing systems, which can be compared and contrasted with natural intelligence. AI2, on the other hand, is a field of study that deals with the design, construction and evaluation of AI1 systems, i.e. artificial systems that manifest intelligence. Because the definition

of AI2 depends on how we understand AI1, which in turn depends on how we understand intelligence itself, defining AI depends on precisely defining intelligence as such. This is problematic because only in the field of psychology this term is considered controversial and there is no consensus on a single definition (as many as 28 new ones have been proposed in the previous decade).[23]

Despite the difficulties in defining AI, it is easier to identify AI examples, referred to as rational agents—systems that receive percepts from the environment and perform actions.[24] The agents act as "intelligent tools," and many of them operate under marketing names (e.g., virtual assistants: Amazon's Alexa, Apple's Siri). They are driven by various types of algorithms (e.g., search, machine learning, evolutionary, artificial neural networks), and when combined with a physical body they become examples of "embodied" AI (e.g., self-driving cars, robots). Human beings, when interacting with AI-based systems, intentionally or unknowingly create hybrid systems.[25] The degree of fusion with artificial entities can be described on a continuum of cyborgisation: from interaction with static (PC), mobile (smartphone) and wearable technologies (smart-glasses), to augmentation (fusing artifacts with the human nervous system).[26] This fusion can be explicit, as in the case of human-cobot systems in the production process, but it can also be implicit to the user of the technology. An example of this are the algorithms that control the mathematical and statistical representation of each Internet user, which, according to Deleuze, can be described as "dividual."[27] This bank of data is created by the activity of the Internet user, but his "mind" is made up of algorithms beyond his control and suggesting customised content. Reacting to it makes the human being (individual) and the "dividual" function in a continuous feedback loop, providing data and reacting to it. They unknowingly meld together to form a kind of augmented mind,[28] which can be referred to in a working way as the "hybrid self."

---

[23] See Dagmar M o n e t t and Colin W.P. L e w i s, "Getting Clarity by Defining Artificial Intelligence: A Survey," in *Philosophy and Theory of Artificial Intelligence 2017*, ed. Vincent C. Müller (Berlin: Springer, 2018), 212–14.

[24] Stuart J. R u s s e l l and Peter N o r v i g, *Artificial Intelligence: A Modern Approach* (Boston: Pearson, 2020).

[25] See Wulf L o h and Janina L o h, "Autonomy and Responsibility in Hybrid Systems: The Example of Autonomous Cars," in *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*, eds. Patrick Lin, Keith Abney and Ryan Jenkins (New York: Oxford University Press, 2017), 35–50.

[26] Alex J u p i t e r, "The Human-Cyborg Continuum: Why AI Is Pointless and Why We Should All Become Cyborgs Instead," June 4, 2016, https://medium.com/@AlexJupiter/the-human-cyborg-continuum-why-ai-is-pointless-and-why-we-should-all-become-cyborgs-instead-4de0c4bb476f.

[27] See Gilles D e l e u z e, "Postscript on the Societies of Control," *October* 59 (1992): 3–7.

[28] See Andy C l a r k and David C h a l m e r s, "The Extended Mind," *Analysis* 58, no. 1(1998): 7–19.

Later in this paper, the results of a study on the assignment of the MS to AI-driven artifacts will be presented. Due to the fact that this process is informal, intuitive reasoning, it is worth looking at how AI is understood by the users themselves. The surveys conducted in seven countries (e.g., USA, Germany, China) show that public awareness of AI seems to depend on the visibility of its use.[29] It was found that 90% of respondents were aware that a voice assistant (visible AI) was based on AI, while only one in three respondents associated online shopping websites, video streaming services and social media (unvisible AI) with AI. The obtained results correspond to the results of studies, which captured differences in the understanding of AI by experts (IT specialists) and laymen.[30] As it turns out, for people with expert knowledge, AI is primarily "algorithmic systems" (e.g. image generation algorithm), while for laymen, it is mainly "nature imitating systems" (e.g. humanoid robot). When categorising AI examples, experts are mainly guided by functional features, while laymen also consider structural features of the systems. The functions of "algorithmic systems" are cognitive, related to performing the so-called objective tasks (e.g., pattern recognition), while "nature imitating systems" perform tasks that seem subjective in nature (based on emotions and intuition).[31]

The identification of AI with embodied, imitating entities found in nature should be attributed to contact with "AI narratives" present in pop culture, which include "portrayals of any machines (or hybrids, such as cyborgs) to which intelligence has been ascribed, which can include representations under terms such as robots, androids or automata."[32] Some narratives are non-fictional (e.g., TV news) and some are fictional (e.g., sci-fi films). In non-fiction AI narratives, attention is paid mainly to the examples of "weak" AI, while the heroes of fiction AI narratives are the examples of AGI. In the latter case, they not only talk and walk, but are capable of feeling human emotions, have elements of self-awareness and free will. In addition, they are characterised by exaggerated corporeality (e.g., T-800 in *Terminator*, 1984) and hypersexuality (e.g., Ava in *Ex Machina*, 2015), they have superhuman resistance to pain and indestructibility.[33]

---

[29] Jem D a v i e s, "AI Today, AI Tomorrow. The Arm 2020 Global AI Survey," armBlueprint, February 3, 2020, https://www.arm.com/resources/report/ai-today-ai-tomorrow-ty.

[30] See Paweł F o r t u n a and Oleg G o r ba n i u k, "What Is Behind the Buzzword for Experts and Laymen: Representation of 'Artificial Intelligence' in the IT-professionals' and Non-professionals' Minds," *Europe's Journal of Psychology*" 8, no. 2 (2022): 207–18.

[31] See Yoel I n b a r, Jeremy C o n e, and Thomas G i l o v i c h, "People's Intuitions about Intuitive Insight and Intuitive Choice," *Journal of Personality and Social Psychology* 99, no. 2 (2010): 232–47.

[32] Stephen C a v e, Kanta D i h a l, and Sarah D i l l o n, *AI Narratives: A History of Imaginative Thinking about Intelligent Machines* (Oxford: Oxford University Press, 2020), 5.

[33] D a v i e s, "AI Today, AI Tomorrow: The Arm 2020 Global AI Survey."