Questioning our Understanding of Intelligence in AI

Cassandra Beyer

The London School of Economics and Political Science

Academic Essay

May 2nd, 2025

Questioning our Understanding of Intelligence in AI

This essay explores the question of whether artificial systems can be intelligent, more specifically, their potential to develop Artificial General Intelligence (AGI). Rather than offering a decisive answer to whether AGI exists today, my aim is to evaluate which conceptual frameworks for understanding intelligence are best suited for an application to artificial entities. A central claim I will defend is that investigations of AGI should not be framed in terms of human-like intelligence. Approaches to intelligence that rely too heavily on human cognition or embodiment risk mischaracterizing the nature and potential of artificial systems.

Current debates about intelligence in AI are often entangled in conflicting assumptions and definitional disagreements about the nature of intelligence itself. These contrasting understandings become evident in my comparison of three differing approaches to AGI, namely pragmatic-behavioral, cognitive, and embodied kinds. Accounts that dismiss the prospects of AGI commonly do so from an overly anthropocentric viewpoint that is closely tied to the cognitive and biological structures of humans. By drawing on accounts like Mollo's (2024), I aim to highlight the advantages of a pragmatic characterization of AGI. Defining AGI in terms of flexibility, generalizability, and adaptive learning allows to more suitable capture the practically significant aspects of intelligence in the context of AI. While claims like Chemero's (2023) that artificial entities “think differently“ may hold true, I argue that such cognitive differences do not sufficiently demonstrate their inability of acquiring AGI overall.

Furthermore, I aim to challenge accounts that refer to a lack of embodiment to deny prospects of AI achieving general, contextual understanding. Instead, I emphasize the commonly underestimated importance social embeddedness has for the acquisition of “true“ understanding.

While LLMs may not yet be fully embedded in our forms of life, their increased immersion in human culture suggests a capacity to develop socially grounded understanding—and thereby general intelligence—over time.

State of the art

In recent years, the most impressive advancements in AI systems have arguably occurred in the development of large-scale machine-learning models, particularly Large Language Models (LLMs). Systems like Open AIs GPT4 employ deep neural networks trained on trillions of data parameters from web data and literature. Through a process known as self-supervised learning, these models are capable of predicting missing pieces of text, thereby gradually improving proficiency in human language use (Mitchell & Krakauer, 2023). As the scale of available training data rapidly increases, LLMs today can exhibit seemingly human-like conversationalist abilities and remarkable problem-solving abilities.

Despite these impressive developments, most existing AI systems are considered to be forms of *narrow AI*. According to Ray Kurzweil (2005), narrow AI refers to systems capable of performing specific tasks within well-defined domains. They are considered “narrow“ as their abilities are limited to designated functions. What seems to be missing is a capacity to generalize knowledge to unrelated contexts – an ability sometimes described as “transfer learning“ (Taylor & Kuhlman, 2008).

The seemingly lacking adaptability of narrow AI to novel contexts marks the main contrast to concepts of Artificial General Intelligence (AGI). Although scholars' optimism about the prospects of AGI differs, it is broadly agreed that AGI today remains hypothetical. If intelligence is spoken off in a narrow sense, LLMs indeed appear intelligent. For my discussion, however, only intelligence in the sense of AGI will be relevant. Therefore, any further use of the term intelligent is intended to refer to a general understanding. The question remaining is how AGI can be conceptualized in the first place.

Artificial General Intelligence

While there is no universally applied definition of AGI, there are a few commonly agreed features that I will base my understanding on. According to Legg and Hunter (2007), these include (i) generalizability across tasks and contexts, (ii) behavioral flexibility, (iii) goal-directedness, and (iv) adaptivity through learning from past interactions. Although these characteristics are widely recognized, scholars' views on how AGI can be achieved and attributed to entities differ. As the full spectrum of proposals can not be done justice at this point, I will limit my analysis to three differing approaches following a distinction by Goertzel (2014):

1. Pragmatic characterizations (Mollo, 2024)
2. Cognitive characterizations (Fridland, 2015; Rowland, 2009)
3. Embodied characterizations (Brooks, 2002; Mitchell&Krakauer, 2023; Pfeifer &Bongard, 2007)

Among these, a primary contrast lies between embodied and cognitive understandings versus pragmatic, behavioral ones. A broadly shared commonality within cognitive and embodied-focused approaches, is that AGI is closely framed in anology to human intelligence and cognition. Chemero (2023) for instance, though not offering a formal definition, states that embodiment and sensory-motory interaction of humans are central to their intelligence. To achieve AGI, artificial entities would not just need complex problem-solving abilities but need embodied existence with needs and motivations. Similarly, Mitchell and Krakauer (2023) claim that intelligence arises from embodied, interactive experiences in the world. Artificial systems lack the situational grounding humans have and are, therefore, unable to attain “true“ understanding or functional language abilities.

Another way to describe this limitation is to speak of *formal* and *informal* knowledge domains. Dreyfus's (1995) essay on "What Computers Can't Do" famously argued that computers only succeed in formal domains that are rule-based and structured. Skills within informal domains, however, require tacit knowledge and practical experiences – a learning mechanism that Dreyfus considered to be out of reach for machines. Similarly, Searle (1980) distinguished between syntax and semantics, claiming that while LLMs may apply language correctly (*syntax*), they lack genuine understanding of meaning (*semantics*).

The underlying idea across these accounts is that the ability to generalize and adaptively apply knowledge - a key feature of AGI - requires "understanding". Embodied and cognitive theories differ in their concrete explanations of how such true understanding is acquired. While embodied theories, like Brooks (2002), ground understanding in physical interaction with the world, cognitive accounts posit internal mental models with representational content (Rowlands, 2009). In either case, scholars' views on how understanding is achieved are closely tied to human-like structures. In contrast, a pragmatic, behavioral conception of AGI does not require human-like understanding. Mollo (2024), for example, views intelligence as a broader, observable phenomenon manifested in systems' flexible and adaptive behavior. This broadly resembles Turing's (2009) view of thinking as a performance capacity that can be objectively measured. Applying a behavioral approach to AGI offers several key advantages highlighted in the following.

AGI vs. Human-like intelligence

Firstly, defining intelligence alongside cognition is problematic in that it requires a definitive characterization of cognition - an equally ambiguous and controversial concept. Within a Philosophy of Mind, theories about the nature of cognition remain heavily debated and vary widely from computational to extended models of cognition (Clark & Chalmers, 1998). As a consensus is unlikely to emerge soon, characterizing intelligence based on human cognition is bound to be counterproductive. In doing so, the question of AGI becomes conflated with one about human-like intelligence. The epistemic distinctiveness and usefulness of the very concept of AGI for philosophical and normative applications within AI ethics becomes lost. Whether AI can be intelligent might as well be a question of whether they are cognitive, which defeats the purpose of establishing a distinct notion of AGI as such.

Depending on the context, certain applications of the concept of general intelligence are more useful than others. If Psychology is the relevant field of interest, framing general intelligence via concepts like psychometric intelligence or "g-factors" appears appropriate. For applications toward artificial entities, however, an overly anthropocentric and origin-dependent characterization of intelligence risks prematurely excluding non-human entities. Similar to Mollo (2024), I consider a broader view of cognition that positions general intelligence as one of its distinctive domains to be more useful. Such a view makes it possible to both acknowledge the prospect for intelligence in AI while still recognizing that the processes underlying AGI differ from human-like cognition.

However, unlike Mollo's (2024) account, I do not find that applying a pragmatic understanding of AGI requires the importance of true, contextual understanding to be disregarded. It is still plausible that the ability of artificial systems to exhibit flexible, adaptive behaviors is made possible by the acquisition contextual understanding. However I find it problematic to assume that such acquisition requires human-like biological or cognitive structures.

Does Understanding require embodiment?

Accounts like Chemeros (2023) or Mitchell and Kraukauers (2023) tend to overestimate the importance of physical, biological embodiment as a prerequisite for informal understanding. Mitchell and Krakauer (2023) for example, claim that LLMs can never understand the word "tickle" because they can never experience its physical sensation. When this argument is applied to other contexts, however, the claim that understanding depends on sensory experiences appears absurd. A paralyzed wheelchair user may never experience the feeling of what it means to be "tickeled". Yet it is unclear why this should prevent them in their ability to gain a common-sense understanding of the application of the term.

Rather than being a matter of shared biological embodiment, our ability to gain contextual understandings of the meaning of words is a matter of being immersed in shared socio-cultural contexts. Or, as Wittgenstein (2009) describes it, "forms of life". Being a member of a certain form of life, allows for a process of linguistic socialization to take place in which we gradually acquire a nuanced understanding of concepts like irony or humor. Through social encounters and interactions, a shared knowledge about the appropriate use of words develops. Physical experiences of sensations like "tickle" may enable more intuitive, deeper levels of understanding but are not an obvious prerequisite. Collins's (1996) account gives further examples to illustrate this: A pet may be able to physically navigate and experience the human world, yet it does not understand the concept "tidy" because it lacks the social context of its meaning: “Immersion in human-like physical and social situations is not sufficient to produce socialization, even for entities whos brain and body are similar to humans“ (1996, p.105).

Similarly, an entity can become part of a social collective and share a form of life without having the same embodiment or physical experiences. If informal understanding is a product of social embeddedness and is not reliant on specific forms of embodiment, then there is no obvious reason to exclude artificial entities from acquiring it. Especially if humanoid AI serving as social companions are considered, it is plausible that a nuanced understanding from their immersions in shared sociocultural contexts can develop.

Conclusion

There is an understandable hesitance against the idea of attributing artificial systems human-like intelligence. However, much of this resistance stems from problematic equations of AGI with human-like cognition, which some even lead to conclude that “LLMs could be slightly conscious“ (Perez, 2022). This is especially troublesome for applications in normative contexts, where cognitive characterizations of AGI can fuel unnecessary concerns about human-like moral status attributions to artificial entities. If a careful distinction between human-like intelligence, cognition and AGI is maintained however, such implications are avoided.

In this essay, I have challenged scholars' emphasis on embodiment and sensory experience for the attainment of informal understandings and highlighted the importance of a shared form of life instead. Despite the plausibility of acquisitions of contextual knowledge through social embeddedness, a certain degree of epistemic uncertainty about the "true" understanding of AI remains. However, this same epistemic problem occurs regarding the existence of consciousness or intelligence in other humans. We might after-all be surrounded by “philosophical zombies“ – a term Chalmers (1997) applies to describe humans that are indistinguishable from us but lack mental states. Given the increased urgency our understanding of intelligence has for normative contexts, philosophical skepticisms of this kind appear hardly justifiable. From a normative perspective, a behavioral approach to intelligence in Mollo's (2024) sense, coupled with AIs level of immersion in our form of life, ends up being the most practically applicable.

What exactly the benchmark for the necessary level of social embeddedness is, will remain open for now. Todays LLMs are likely still far from being embedded in our forms of life to the required degree that enables their linguistic socialization. Even if AI’s ability for contextual knowledge applications improves, we should remain cautious in attributing AGI too hastily and solely based on technological advancements. Keeping the possibility for linguistic socialization in mind however, allows to recognize the prospects for intelligence in AI in a more meaningful, non-anthropocentric manner.

References

Burkart, J. M., Schubiger, M. N., & van Schaik, C. P. (2017). The evolution of general intelligence. *Behavioral and Brain Sciences, 40*, e195. https://doi.org/10.1017/S0140525X16000959

Chalmers, D. J. (1997). *The conscious mind: In search of a fundamental theory*. Oxford Paperbacks.

Chalmers, D. J. (2014). Intuitions in philosophy: A minimal defense. *Philosophical Studies, 171*(3), 535–544.

Chemero, A. (2023). LLMs differ from human cognition because they are not embodied. *Nature Human Behaviour, 7*(11), 1828–1829.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis, 58*(1), 7–19. Coelho Mollo, D. (2024). Intelligent behaviour. *Erkenntnis, 89*(2), 705–721.

Collins, H. M. (1996). Embedded or embodied? A review of Hubert Dreyfus' *What Computers Still Can't Do*. *Artificial Intelligence, 80*(1), 99–118.

Dreyfus, H. L. (1992). *What computers still can't do: A critique of artificial reason*. MIT Press.

Fridland, E. (2015). Learning our way to intelligence: Reflections on Dennett and appropriateness. In C. Muñoz-Suárez & F. De Brigard (Eds.), *Content and consciousness revisited* (Vol. 7, pp. 129–148). Springer. https://doi.org/10.1007/978-3-319-17374-0_8.

Larson, E. J. (2021). *The myth of artificial intelligence: Why computers can't think the way we do*. Harvard University Press.

Mitchell, M., & Krakauer, D. C. (2023). The debate over understanding in AI's large language models. *Proceedings of the National Academy of Sciences, 120*(13), e2215907120.

Perez, E., Ringer, S., Lukošiūtė, K., Nguyen, K., Chen, E., Heiner, S., Pettit, C., et al. (2022). Discovering language model behaviors with model-written evaluations. *arXiv preprint* arXiv:2212.09251. https://arxiv.org/abs/2212.09251

Pfeifer, R., & Bongard, J. (2007). *How the body shapes the way we think: A new view of intelligence*. Boston Review.

Rowlands, M. (2009). Extended cognition and the mark of the cognitive. *Philosophical Psychology, 22*(1), 1–19.

Searle, J. R. (1980). Minds, brains, and programs. Behavioral and brain sciences, 3(3), 417-424.

Taylor, M. E., Kuhlmann, G., & Stone, P. (2008). Transfer learning and intelligence: An argument and approach. *Frontiers in Artificial Intelligence and Applications, 171*, 326–331.

Turing, A. M. (2009). Computing machinery and intelligence. In *Parsing the Turing Test* (pp. 23–65). Springer Netherlands.

Wittgenstein, L. (2009). *Philosophical investigations* (Rev. 4th ed.). Wiley-Blackwell.