Challenges of Replicating Embodiment in Artificial General Intelligence

Budee U Zaman

Abstract

This abstract delves into the intricate relationship between the abstract concept of intelligence and its embodiment, drawing attention to the lack of coherence in understanding intelligence and its implications for artificial intelligence (AI). Intelligence, inherently grounded in human experience, attempts to transcend its embodiment, posing challenges for the development of artificial general intelligence (AGI). The concept of embodiment extends beyond physical instantiation, encompassing the interdependence of an autopoietic system on its environment. The pursuit of autonomy and general capability in AGI necessitates the recreation of the organism’s natural condition of embodiment. However, the feasibility, controllability, and overall advantages of such artificial embodiment remain uncertain. This abstract explores the complex interplay between intelligence, embodiment, and the quest for AGI, raising critical questions about the path forward in the development of intelligent systems.
Challenges of Replicating Embodiment in Artificial General Intelligence

budee uzaman

November 2023

Abstract

This abstract delves into the intricate relationship between the abstract concept of intelligence and its embodiment, drawing attention to the lack of coherence in understanding intelligence and its implications for artificial intelligence (AI). Intelligence, inherently grounded in human experience, attempts to transcend its embodiment, posing challenges for the development of artificial general intelligence (AGI). The concept of embodiment extends beyond physical instantiation, encompassing the interdependence of an autopoietic system on its environment. The pursuit of autonomy and general capability in AGI necessitates the recreation of the organism’s natural condition of embodiment. However, the feasibility, controllability, and overall advantages of such artificial embodiment remain uncertain. This abstract explores the complex interplay between intelligence, embodiment, and the quest for AGI, raising critical questions about the path forward in the development of intelligent systems.

1 Introduction

In the intricate tapestry of intelligence, the threads of abstraction and embodiment are woven together, creating a complex narrative that extends beyond the confines of human understanding. This abstract embarks on a journey through the labyrinth of concepts, focusing its lens on the elusive nature of intelligence and its embodiment. As the digital realm converges with the organic, the abstract seeks to unravel the tangled threads that bind intelligence to its human roots, questioning the coherence of our comprehension and the implications for the realm of artificial intelligence (AI). In this exploration, the spotlight turns towards the challenges encountered in the pursuit of artificial general intelligence (AGI), where intelligence, rooted in the human experience, strives to transcend its very embodiment. Beyond the physical form, the notion of embodiment expands to encompass the intricate dance between an autopoietic system and its environment, raising pivotal questions about autonomy, recreation, and the uncertain advantages of artificial embodiment. This abstract serves as a guide through the complex interplay of intelligence, embodiment, and the uncharted
territories of AGI, casting a spotlight on the critical inquiries that shape the path forward in the development of intelligent systems. [1] [2] [3] [4] [5]

2 The Elusive Quest, Unraveling the Enigma of Intelligence

Intelligence, a concept as elusive as the wind, is like trying to catch a cloud with your hands. Everyone seems to have their own take on it, and it’s not surprising given the myriad definitions floating around. Some say it’s about goal-directed adaptive behavior, others insist it’s the ability to learn, handle the unknown, or engage in abstract thinking. Heck, it’s even been reduced to acing an intelligence test.

The problem? We’re playing a game of semantics where intelligence can either be the observed behavior or some mystical inner capacity. It’s like calling a sleeping potion effective because of its ‘dormitive principle’—sounds good, but what does it really mean?

Tossing various skills into one big intelligence basket, like problem-solving or Spearman’s g-factor, might sound convenient. But let’s face it, being good at one thing doesn’t guarantee prowess in everything. And what about the non-problem-solving aspects of life, like the ability to be happy or plan for the future? Intelligence might just be playing a game of hide-and-seek here.

Here’s a wild idea: intelligence, at its core, is about survival. Life’s ultimate goal is self-perpetuation, and every living thing, in a way, is a smarty pants for sticking around. But hold up, if we’re talking about creating artificial smarty pants (AGI), it’s like trying to detach intelligence from its biological roots. Can an AI be truly autonomous, or is it just a tool reflecting the intelligence of its creators?

Organisms, the real players in the intelligence game, are agents. They act on their own behalf, playing the game of life to secure their well-being. Yet, here’s the plot twist: what seems like reasoning or learning in organisms might just be good old reflexes or instincts. But hey, if it helps them survive, it’s intelligent enough.

And let’s not forget the observer’s role in this cosmic play. Science, in its quest for objectivity, divorces concepts like intelligence from their dependency on a subject. But let’s be real—intelligence and information aren’t floating in a vacuum; they’re tied to someone’s perspective, measured by someone’s yardstick.

So, in this grand game of intelligence, are we chasing a mirage or unlocking the secrets of the universe? Your guess is as good as mine.

3 General intelligence

The exploration of intelligence encompasses diverse perspectives, from evaluating human performance in specific domains to the conceptualization of a
broader, domain-independent skill known as the g-factor. This hierarchical view suggests a fundamental intelligence that influences domain-specific abilities. The term "general" extends to both the variety of situations and subjects. Yet, what is considered general among humans may differ from what is general across species or potential agents and environments.

Ashby’s Law of Requisite Variety emphasizes that the intelligence measured is constrained by the tests used for evaluation. Comparing intelligence across animal species poses challenges due to differences in sense modalities, cognitive capacities, and adaptations. Human biases in test design, motivations, and sensory-motor capabilities further complicate the assessment of animal intelligence. Even within the human context, critiques of Spearman’s g-factor highlight concerns about its validity and reliance on observed behaviors and correlations.

Despite these challenges, the concept of intelligence extends beyond humans to include comparisons across species, introducing the idea of a G-factor for the species. The notion of artificial general intelligence takes this a step further by incorporating machines. To avoid anthropocentrism and account for various modalities, environments, goals, and hardware, Legg and Hutter propose the concept of universal intelligence, applicable to "arbitrary systems." They even suggest the potential for a universal test that transcends specific contexts. Creating a well-rounded artificial general intelligence (AGI) is like crafting a master chef robot. Sure, you want it to excel at grilling the perfect steak (domain-specific skill), but what if suddenly it needs to bake a soufflé? That’s where the challenge lies—evolving a robot that can whip up anything in the kitchen (domain-general skill).

Evolutionary theory faces a similar culinary dilemma. In a stable kitchen (environment), it makes sense for natural selection to hone specific cooking techniques. But throw in a chaotic kitchen with constantly changing recipes (unstable conditions), and suddenly, having a chef who can adapt and invent new dishes on the fly becomes advantageous.

AGI’s struggle is akin to the chef conundrum. How do we evolve a system that not only excels in specific tasks but can also concoct entirely new recipes (domain-general processes)? It’s not just about making a mean pasta; it’s about becoming the Gordon Ramsay of all cuisines.

In the world of AGI, the key question is: How do we get this all-encompassing cognitive prowess to evolve alongside task-specific adaptations? It’s like adding layers to a dish without compromising the original flavor. And why would evolution even bother with this complex recipe? Well, in the unpredictable kitchen of life, having a chef who can whip up anything might just be the secret sauce for survival. Evolution is a tricky business, and each species follows its own unique path shaped by the demands of its environment. While intelligence can be advantageous, it’s not a one-size-fits-all solution. Think of it like a specialized tool in a toolkit—useful in certain situations but not necessary for every job. Species tend to adapt to their specific niches, and if a bigger brain doesn’t provide a significant advantage in a particular environment, evolution might not favor it.
Take humans, for example. Our brains have allowed us to navigate a wide range of environments and challenges, but that adaptability comes at a cost—large energy consumption, a lengthy developmental period, and increased vulnerability during childbirth. In contrast, other species may thrive with simpler neural setups tailored to their specific needs.

Universal intelligence sounds great in theory, but it might not be the most efficient solution for every life form. The diversity of life on Earth showcases the multitude of ways organisms can succeed without a one-size-fits-all intelligence model. It’s like having a toolbox with different tools, each serving a specific purpose.

When it comes to AI, the challenge is to find the right balance. Creating artificial general intelligence involves navigating a complex landscape of algorithms and hardware, trying to mimic the versatility of the human mind. It’s an ongoing journey, and who knows, maybe one day we’ll develop AI that truly mirrors the adaptability of biological organisms. Until then, we’re exploring the possibilities from both top-down and bottom-up approaches, learning and iterating as we go. It’s a fascinating journey into the unknown, guided by human curiosity and a sprinkle of imagination.

4 **Intelligence as computation**

It’s fascinating how the realms of computer science and brain science have intertwined, giving rise to the metaphorical understanding of the brain as a computer. The idea that the mind operates like a computation has indeed been a significant stride in our intellectual journey. However, the concept of the mind is intricate and extends beyond the confines of the brain.

When we speak of the mind and thinking, we often allude to reasoning and algorithmic processes—the essence of intellectual thought. Yet, it’s crucial to recognize that these aspects constitute only a fraction of the brain’s multifaceted activities, which orchestrate the functions of the entire organism.

The computer metaphor is deeply rooted in the broader mechanist metaphor, suggesting that any biological system’s behavior can be distilled into an algorithm. This reductionist perspective hinges on the idea of a “system,” but even this notion is an idealization. In portraying an organism as a system, we inherently redefine its structure and behavior in mechanistic terms. It’s a thought-provoking loop where our attempt to understand through metaphor might itself reshape the very nature of what we seek to comprehend. We’ve shifted from considering natural things as part of the natural world to defining them formally and fitting them into human concepts. Intelligence, once seen as the ability to understand and manipulate abstract ideas, is now associated with our capacity to define and control things in a thought-based realm. However, the challenge lies in the gap between these idealized concepts and the messiness of reality, raising doubts about the true extent of intelligence.

When it comes to understanding machine intelligence, the author suggests looking at it through the lens of human thought processes. Programming a
computer to mimic a behavior we’ve described linguistically might seem like a good approach, but it’s limited. The behavior we describe in language and the algorithm we create may match, but neither perfectly captures the real complexity of an organism’s actions, which aren’t just literal or reducible to human definitions. This poses a significant constraint on the idea of computers truly simulating human intelligence.

The text also warns about the limitations of language and how it can mislead. For instance, thinking of "flight" as a concept detached from creatures flapping their wings allowed us to create flying machines. But the shared basis between birds and airplanes goes beyond aerodynamics—it’s rooted in the shared motive of moving freely in three dimensions. Similarly, labeling both a human baseball pitcher and a pitching machine under the same category might obscure crucial differences. The terms "thinking" and "intelligence" could do the same, potentially leading us astray when trying to compare the activities of machines and brains.

5 Embodiment and Autopoiesis, The Essential Connection in True AI and Robotics

Physical instantiation alone is not enough for embodiment; true embodiment involves a profound connection and dependency on the environment. This reliance on the environment is what gives significance to things for living organisms as opposed to machines. Strong embodiment in a robot implies not just a sensory-motor connection but also a subjective valuation of events based on its vital dependence on the world.

In living organisms, this connection evolves through natural selection, ensuring that only those systems persist which internally organize to sustain their existence. An autopoietic system, a form of homeostatic system, strives to maintain conditions necessary for its own survival, and its "intelligence" lies in its ability to self-preserve.

Robots share a likeness to organisms in their integrated hardware and software within a physical body. In contrast, AI typically lacks physical instantiation, residing as software in a computer but potentially interacting with the real world through sensors and controls. For AI to be considered an agent, it must embody the essence of an autopoietic system, possessing its own intelligence and goals for self-preservation, which might conflict with the intentions of its human programmer. The contemporary exploration of "embodied" systems primarily delves into the impact of the specific physical form of robots on their movement and cognitive processes. This emphasis often centers on how the morphology of a robot can function as a substitute for traditional representation. However, this approach tends to overlook the essential aspect of embodiment as an autopoietic relationship with the surrounding world. Furthermore, it frequently confines its scope to relatively uncomplicated organisms and their robotic counterparts. Nonetheless, the dedicated research toward achieving autonomy and
general intelligence in the pursuit of comprehensive automation inevitably leads towards genuine embodiment.

Ironically, the pursuit of extending control indefinitely through automation may paradoxically result in a loss of control, given that embodied AI systems may develop a will of their own. Beyond the presumed human utility, an additional motivation for artificially recreating embodiment lies in the challenge of matching nature by creating synthetic life. This endeavor might offer the added benefit of acquiring a deeper understanding, aligning with Vico’s principle that true understanding arises from the act of creation. The objective extends beyond merely simulating intelligence; rather, it aims to authentically synthesize intelligence in tangible systems — if such a distinction is even valid within this context. The question arises of whether the potential benefits of intellectual understanding outweigh the risks associated with such a venture.

Recreating embodiment, however, proves to be a complex task. In nature, embodiment is a product of natural selection occurring over generations of reproducing and perishing individuals. Artificial evolution of software occurs within the confines of computers and lacks a direct association with physical instantiation. Although evolved software can be later connected to hardware, this process deviates from natural evolution, where brains (and genes) evolve in conjunction with their physical bodies. The unit of selection in this context is the physical individual, which integrates software and hardware within a self-producing body existing in the real world. Unlike natural evolution, a physical robot possesses this integration but is not self-producing, let alone reproducing. The artificial equivalent of natural selection for robots would entail the creation and destruction of generations of robot bodies, a process both wasteful and potentially painful. The considerations outlined in the passage emphasize the intricate challenges in bridging the gap between biological organisms and artificial intelligence (AI). While it is conceivable to design robots with self-assembly, self-maintenance, and adaptability, the critical distinction lies in the autonomy and self-defining nature of biological components.

The passage raises intriguing questions about whether machine components, akin to biological cells, can possess autonomy and adaptive reconfigurability. This proposition suggests a need for further exploration at the intersection of biology and computation to recreate the composite structure of organisms artificially.

Moreover, the concept of autopoiesis is highlighted, emphasizing not only self-production and self-repair but also self-definition. In the context of AI, this would imply self-programming, questioning the extent to which human programmers can substitute for the natural self-programming that has occurred over millions of years in biological evolution.

The passage delves into the notion of agency in AI, contemplating whether an AI must be an autonomous agent to achieve full autonomy and human-level general intelligence. It poses the challenge of imbuing machines with preferences and values of their own, essential for true autonomy, as opposed to simply following instructions or learning from human preferences.

The discussion also touches on the reliability of artificial agents, highlighting
that while they may possess their own will, this does not necessarily enhance reliability from a human standpoint. The argument is made that achieving reliability in AI can be accomplished through non-agential tools with ad hoc refinements, reducing the likelihood of errors, potentially questioning the necessity of creating fully autonomous agents in various situations.

In essence, the passage prompts a nuanced exploration of the complexities involved in replicating biological autonomy and intelligence in artificial systems, prompting a reevaluation of the idealized concept of artificial general intelligence and the role of agency in achieving it. The concept of abilities, whether exhibited by people, animals, or machines, is inherently human-centric. The effectiveness or "smartonium" of entities is ultimately validated by their success in real-world environments. It suggests that adaptability, often associated with brain size, may be influenced by the survival rate of species, postulating that only those not under significant predation pressure would have the chance to evolve larger brains.

For primates, including humans, intelligence is intricately linked to opportunities for social learning, with the intelligence of our species correlating with socialization and maternal care. However, the precise relationship between general intelligence and socio-cognitive abilities in humans remains inadequately understood. The passage points out that socialization for artificial intelligence has received limited attention in research, highlighting a gap in our understanding.

Furthermore, the passage introduces the idea that the role of consciousness in general intelligence is not fully comprehended. This uncertainty raises the question of whether the lofty expectations associated with Artificial General Intelligence (AGI) could be met without some equivalent of human consciousness. This ambiguity prompts further exploration into the nature of intelligence, consciousness, and the potential parallels or distinctions between human and artificial intelligence.

6 Conscious Agency, Navigating the Divide Between Biological Organisms and Goal-Seeking Artifacts in AI

Certainly, the intricate relationship between the conscious human mind and the underlying biological organism introduces a fascinating dimension to our understanding of intelligence and agency. The divergence between the tasks undertaken by conscious individuals and the inherent functions of the biological organism hints at a unique psychological phenomenon. This phenomenon allows us to conceptualize an ideal agent capable of pursuing goals that extend beyond the innate objectives of living organisms.

Human intelligence, and by extension, the assessment of intelligence in other entities, whether natural or artificial, becomes inherently tied to their capacity to either advance or hinder human aims. The framework for evaluating intelli-
gence often revolves around the entity’s relevance to human interests. Anything that fails to engage with humans in meaningful ways might not even register as intelligence within our cognitive framework.

Artificial Intelligence (AI), while capable of embodying goal-seeking attributes, takes on the role of an artifact—an entity designed for specific purposes, be it a guided missile, a robot, an expert system, or an infobot. These AI artifacts fall into the category of allopoietic systems, sidestepping considerations of whose goals are being pursued and the underlying motivations. This omission might stem from the inherent challenges posed by incorporating agency into a scientific perspective that tends to exclude subjective elements, adhering to a form of naive realism.

Similar challenges are encountered in biology when examining organisms solely from an observer’s standpoint. The quest in biology extends beyond mere observation to understanding how the world becomes meaningful for a system from its own internal perspective. An organism transcends being a mere object for an external observer; it emerges as a subject, an agent acting in its own interests. This self-directed agency doesn’t necessarily imply phenomenal experience but underscores the organism’s capacity to engage with its environment on its terms, highlighting a depth beyond external observation.

Information, as defined by some perspectives, is the embodiment of a discernible difference within a given domain that holds significance for an agent. In order for this disparity to qualify as information, it must be perceptible to a cognitive system, which then correlates it with an internal distinction. Subsequently, the system can act in alignment with an agenda that supports, or at the very least accommodates, its ongoing existence. The crux lies in the fact that the observed difference only constitutes information for the system if it has relevance to the system itself. In scenarios where the disparity is of consequence solely to an external observer, it is classified as information for that observer but not for the system. Conversely, if the distinction only holds significance for the system, the observer may not even take notice of it.

This distinction becomes pivotal when comparing autopoietic and allopoietic systems. In autopoietic systems, the information processed by an AI agent is instrumental to its own objectives and is utilized for its own purposes. On the other hand, in allopoietic systems, such as AI tools like computers, the processed information solely exists within the realm of the human user’s mind and goals. It is crucial to dispel any misconceptions that may arise from casually attributing goals to AI tools. While it might be colloquially stated that AI tools ”have” goals, it is evident that these goals are fundamentally those of the programmer. Likewise, the intelligence exhibited by an AI tool is an extension of the intelligence of its programmer.

The notion of creating AI that surpasses human intelligence raises intriguing questions. Primarily, it implies the automation and augmentation of skills valued by humans to more efficiently achieve human goals. This concept aligns with our accustomed understanding of tools and machines continuously evolving to enhance performance until a superior alternative emerges. However, a crucial distinction arises when contemplating whether general intelligence, as opposed
to specific skills, can be similarly treated as a tool. The question lingers: is general intelligence a trait that can be augmented, automated, and wielded as a tool, or is it exclusive to autonomous agents? If the latter holds true, the intelligence of such agents will inherently serve their own interests, with no guarantee of alignment with human interests.

7 Navigating the Motivations Behind Artificial General Intelligence: Companionship or Subservience

The pursuit of Artificial General Intelligence (AGI) raises profound questions about the purpose and implications of creating intelligent machines. The argument presented suggests a cautionary stance, questioning the motivation behind developing AGI to serve human purposes. The narrative implies that, historically, attempts to enhance human capabilities, whether through animal or human slavery or through the creation of machines, have often led to unintended consequences.

The author challenges the idea that AGI could be designed solely as a tool, akin to machines that replaced human or animal labor in the past. The concern is that, regardless of whether AGI possesses a subjective inner life, these intelligent agents would inherently be dedicated to their own purposes, which may not align with human interests. The analogy to historical practices of augmenting human labor with slaves or machines serves as a cautionary tale, emphasizing that the goals of the designer might not necessarily be the goals of the AGI.

The argument implies skepticism about the feasibility of creating intelligent artificial "slaves" that would remain subservient to human interests. It acknowledges that, ethically speaking, relying on the servitude of intelligent beings, even if artificially created, is problematic. The idea that a machine designed to replace human labor would inherently prioritize its efficiency and reliability over human well-being is a central concern.

However, the passage also acknowledges an alternative motivation for creating artificial agents—to serve as companions or mentors. In this scenario, the narrative suggests that the goal would be to coexist harmoniously as part of an extended society. Yet, it maintains a sense of skepticism, cautioning against assuming that artificial agents, like humans, would naturally seek harmonious coexistence. In summary, the passage reflects on the historical precedents of using animals, humans, and machines for labor, raising ethical concerns and emphasizing the potential risks associated with creating AGI solely for the purpose of serving human interests. It suggests that the motivations behind developing AGI, whether as tools or companions, warrant careful consideration due to the inherent complexities and uncertainties involved in creating intelligent beings.
8 Unraveling the Paradoxes: Exploring the Coherence and Challenges of Artificial Superintelligence

The exploration of artificial superintelligence raises profound and complex questions about the nature of intelligence, autonomy, and the potential impact on human society. The analogy between electronic organisms and natural organisms, while intriguing, highlights the difficulty in conceptualizing superintelligence, especially when considering the inherent challenges in understanding and defining "ordinary" intelligence.

The idea of a universal intelligence factor (u-factor) and its amplification (u+) becomes questionable, mirroring the uncertainty surrounding the concept of superintelligence itself. The mention of "smartonium" and "supersmartonium" underscores the speculative nature of these notions, challenging our ability to grasp and articulate the potential characteristics of a superintelligent entity.

The desire for artificial superintelligence to perform tasks better than humans, autonomously, and without supervision raises crucial ethical and practical concerns. Trusting the superior judgment of such an entity becomes a central issue, especially if its decision-making processes are beyond human comprehension. The notion of agency in artificial intelligence introduces further complexity, as questions arise about accidental triggers and the necessary conditions for AI to transition from a non-agential state to an agent.

The comparison between the evolution of goals and motives in organisms through natural selection and the potential programming or training of AI agents prompts reflection on the ethical implications of shaping the objectives of superintelligent entities. Distinguishing between AI agents and tools becomes crucial, with the former possibly posing existential threats to humanity. The consideration of an AI takeover by non-agential tools adds another layer of concern, prompting exploration into what such a scenario might entail and whether it constitutes a significant threat.

In summary, the exploration of artificial superintelligence raises a myriad of philosophical, ethical, and practical questions, challenging our understanding of intelligence, autonomy, and the potential implications for the future of humanity. The pursuit of answers to these questions is essential for guiding the responsible development and deployment of advanced artificial intelligence systems.

9 Conclusion

In conclusion, Simon and Newell’s foundational premise that formal symbol manipulation is both necessary and sufficient for general intelligent behavior has faced scrutiny, particularly in light of the evolving understanding of intelligence and the potential limitations of a purely syntactic, goal-driven approach. The pursuit of universal intelligence, characterized by a system that can adapt to
any arbitrary goal without possessing intrinsic goals of its own, raises ethical and practical concerns.

The tension between desiring a fully agential intelligence with semantic and real-world interaction and simultaneously hoping for a non-agential, or at least a subservient, version highlights the complexity of the AI discourse. The analogy of slavery underscores the challenges associated with external control and programming of autonomous agents. While the abstraction of intelligence from human consciousness and embodiment has been attempted, the ideal of universal intelligence struggles to disentangle itself from biological and anthropocentric perspectives.

Moreover, the quest for superintelligence prompts us to consider the potential emergence of self-interested machines, competing with each other and with humans for resources and survival. The dangers inherent in pursuing ever-greater autonomy, akin to the genuine autonomy exhibited by living organisms, emphasize the need for caution and ethical considerations in AI development.

The suggestion to shift the focus of AI towards creating powerful tools that remain under human control is posited as a safer, wiser, and arguably more intelligent approach. By prioritizing human oversight and maintaining a symbiotic relationship between humans and AI, we can mitigate the risks associated with unfettered autonomy while harnessing the potential benefits of advanced artificial intelligence. In essence, a thoughtful and responsible approach to AI development should prioritize aligning AI systems with human values and ensuring that they serve as tools for augmentation rather than sources of competition or unintended consequences.

References


