

## **The Philosophical Implications of Artificial Intelligence**

### ***How Can the Implementation of AI Technologies Contribute to and Not Degrade Human Flourishing?***

**By Matthew David Segall**

#### **Abstract**

Following the famous 1956 Dartmouth Conference, where a team of computer scientists first coined the term “artificial intelligence,” philosophers have raised significant ontological and ethical questions regarding AI’s nature and its implications for human flourishing. More recently, the growing popularity of Large Language Models and so-called “deep fakes” has made the public more aware of the power of AI, thrusting profound existential dilemmas into our collective consciousness. This chapter attempts to address our current situation from a philosophical perspective by reflecting on basic questions, including: What are these technologies for? Whose interests do they serve? And perhaps most importantly, how are they changing our sense of conscious human agency? The aim is to better contextualize the transformation already underway in the hopes of avoiding false myths and unethical outcomes.

#### **Introduction**

Following the famous 1956 Dartmouth Conference, where a team of computer scientists first coined the term “artificial intelligence,” philosophers have endeavored to raise significant ontological and ethical questions regarding AI’s nature and its implications for human flourishing. In response to the Soviet Union's successful launch of the Sputnik satellite two years after the Dartmouth conference, the US military created what is now known as DARPA (Defense Advanced Research Projects Agency). DARPA quickly became the largest funder of AI research, making clear that, alongside corporate profits, the weaponization of computer technologies has been a primary driver of innovation from the beginning. Today, with companies competing to implement the next generation of AI “agents,” fundamental questions remain regarding the compatibility of such technologies with human flourishing. While the military-industrial complex has been developing these technologies for decades, the recent popularization of AI tools in the form of consumer products has brought the profound ontological and ethical

quandaries they raise into public consciousness. This chapter attempts to address our current situation from a philosophical perspective by reflecting on basic questions, including: What are these technologies for? Whose interests do they serve? And perhaps most importantly, how are they changing our sense of our own conscious human agency? The aim is to better contextualize the transformation already underway in the hopes of dispelling some false myths and avoiding unethical outcomes.

In any discussion about AI, it is crucial from the outset to recognize that human consciousness—our thinking, feeling, and willing—is not made of digital information or reducible to computational processes, no matter how powerful. Throughout history, humanity has attempted to make sense of the mystery of our own minds by making analogies to the popular artifacts of the day: wind-harps, wax tablets, clocks, steam engines, telegraph lines, and today, the computer. While something is always revealed by the use of such metaphors, much may also be obscured. Consciousness is a living medium that remains undetectable by any measuring device that neuroscientists or physicists might devise. While for conscious beings like you and me, our first-person awareness and agency is the most obvious aspect of our existence, its invisibility to any external means of detection highlights a fundamental distinction that must be established at the very start of any discussion of the philosophical significance of AI. Of course, neuroscience utilizes various scanners to track cranial blood flow and electrochemical activity, but such third-person observables are not what we know from within as our conscious experience. You cannot crack open the skull to discover the color red or the taste of a strawberry, much less the source of our moral and aesthetic judgments. How what brain scanners measure might relate to our conscious experiences remains a deep philosophical mystery, with over 200 different proposals currently attempting to address the question and exploring the implications of different answers (Kuhn 2024).

The so-called “hard problem of consciousness,” first formulated by the philosopher David Chalmers (1996), ought to deflate the hype surrounding the ill-formed idea of “conscious” computers. As Chalmers famously argued, scientists could in principle come to understand everything there is to know about the functions of the brain and still have no explanation as to why brains seem to be haunted by consciousness. From a purely physicalist point of view,

consciousness appears to be something “extra” not reducible to the neurochemical goings-on in the skull. Thus, until we have a better philosophical understanding of the nature of consciousness and its place in the physical world, claims that current or future AI devices possess consciousness remain dramatically overblown and even nonsensical.

### **The Embodiment of Consciousness and the Limits of AI**

Understanding the brain itself already requires more than just considering it as an information processor. To truly grasp its nature, we must remember the brain’s embeddedness within the living body and its evolutionary history. Our bodies are ecosystems comprised of trillions of cells, with only about a tenth containing human DNA. The rest are part of a microbial ecosystem symbiotically interacting with our human cells. Thus, we are a living society of “occasions of experience,” to use the philosopher Alfred North Whitehead’s terms (1929), a conscious animal made of a community of other sentient cellular beings—a whole made of smaller wholes, rather than a mere assemblage of algorithmically interacting parts.

Despite the popularity of the metaphor, the brain itself cannot be simplistically divided into mental “software” and physical “hardware.” In the living world of our embodied experience, mind and matter are not so easily separated. While their connection remains mysterious, philosophers like Whitehead argued that we can take a huge step forward in our understanding by coming to see mind and matter as entangled phases in an unbroken creative process. Our minds are embodied processes emergent from ancient evolutionary lineages rooted deep in our cellular architecture. Our capacity for conscious agency depends upon the basic feelings of both sympathy and precarity that typify our existence as fragile membrane-bound creatures metabolically surfing the energy gradients of our environments to survive and thrive. The AI systems currently under development lack anything like this precarious embodiment and historical embeddedness. AI systems are not continually engaged in materially and formally making themselves, as living organisms are. This means, among other things, that they lack any sense of relevance and emotional valence allowing them to determine which data in their environment may be important, and which safely ignored.

The lack of evolutionary embeddedness turns out to be rather important. Machine learning algorithms can acquire certain behaviors in a relatively short training period, but this is not at all the same as how an evolutionary lineage emerges. An evolved organism has been bound up in a process of structural coupling with a whole historical series of environments over the course of billions of years (Maturana and Varela, 1980). Many valuable survival strategies and sense-making heuristics have been picked up along the way—a deep organismic memory of how to be in sync with the rhythms of this planet has been acquired over that long, multi-billion-year process. On the other hand, any machine that is initially designed and built by human beings, even if it is a machine learning neural network trained to interact with specific environments and languages, etc., is still going to lack that depth of embodied memory. Putting their other important differences to the side, no machine has been engaged in evolutionary learning for as long as biological organisms have. This fact gives us reason to suspect that our machines, increasingly impressive though they may be, are relatively blind to important subtle features of the Earth's environment that we, as organisms, are more constitutively prepared for. We should expect even the most sophisticated AI systems to make rather obvious mistakes because they just haven't encountered certain situations before.

### **The Evolutionary Context of Intelligence**

To understand the significance of AI, we must consider intelligence within the broader context of evolution. Evolutionary theory reveals intelligence as not merely the ability to process information but as a deeply embodied skill, shaped by the interaction between organisms and their environments over millions of years. For instance, a bird's ability to navigate long migratory journeys is not merely a matter of calculating distances but is rooted in a complex web of evolutionary adaptations—sensory, neurological, and behavioral—that allow it to adaptively align with magnetic fields, weather patterns, and ecological cues. Human intelligence, likewise, is not just about abstract reasoning or problem-solving: it is about navigating the world as a living, feeling, and historically and culturally situated community member.

AI, as currently conceived, lacks this evolutionary grounding. It may excel at specific tasks, such as playing chess or finding oil deposits, but these abilities do not equate to the kind of general intelligence and sense-making capacity that living organisms possess. Even the most advanced

AI systems do not *understand* the world they operate in; they manipulate numerical weights and statistical patterns without any awareness or understanding of what the digits being processed represent to humans. This distinction is crucial because it underscores the difference between machine learning, which is driven by data and algorithms, and human learning, which is driven by lived experience, emotional investment, and ethical engagement with others and the world.

### **Relevance Realization and the Non-Computational Nature of Cognition**

One of the key insights into the limitations of AI and its implications for human agency comes from recent work in the biology of cognition on “relevance realization.” Jaeger et al. (2024) argue that the ability to realize relevance is observable in all living organisms, from bacteria to humans. However, despite being perfectly natural, organismic relevance realization transcends formalization and so is noncomputable. While computational models may partially simulate some aspects of cognition, they can never fully instantiate this core competency of living beings. In line with what has already been said above, the authors point to crucial metabolic, ecological, and evolutionary dynamics that make organisms dramatically different from any known AI or machine learning systems. These dynamics are involved in the realization of relevance through a meliorative process that allows agents to maintain and optimize their grip on their environments, achieving what they term a “transjective” agent-arena relationship that transcends our normal sense of a boundary separating embodied minds from their environments. Cognition is understood to be radically relational and extended, both socially and ecologically. This relationship, characterized by what the authors call “embodied ecological rationality,” is fundamental to life and categorically distinguishes organisms from non-living computer systems.

From this perspective, the difference between living organisms and AI systems becomes even more pronounced. AI operates within a predefined formalized ontology, handling well-defined problems in a “small world.” In contrast, organisms navigate a “large world” filled with ill-defined problems, where relevance must be continuously realized to make sense of their environment. Jaeger et al.’s distinction between logical inference (i.e., the operations instantiating machine learning algorithms) and relevance realization echoes the Whiteheadian distinction (Whitehead, 1929, p. 199-207, 280) between the Bayesian statistical calculations informing neural network architectures and organismic “lures for feeling,” which while also

capable of predicting probabilities are rooted in aesthetic contrasts and an attunement to long-term environmental rhythms rather than numerical calculations or statistical models.

### **The Nature of Intelligence: Human and Artificial**

It is important to deepen our understanding of the nature of “intelligence.” Even human intelligence could already be considered artificial in some respects, since so much of our everyday awareness and agency is augmented by our tools and artifacts. Our ancestors have been coevolving with external implements for tens of thousands if not millions of years, whether in the form of obsidian fashioned to carve animal meat, bone flutes to play music, or spoken words to coordinate shared attention and intention. Speech externalizes thought in a physical form, a process further extended by written languages, the printing press, the telegram, radio, television, and the Internet. Thus, rather than imagining that some more advanced AI system might become “truly” or “generally” intelligent, digital forms of artificial intelligence should be understood as the latest evolutionary extension of human intelligence, which has always already been, to some degree, artificial.

Considering the irreducibility of consciousness to any measurable physical process and our species’ coevolutionary history with external artifacts, the crucial question is not whether machines will become conscious. AI devices are digital information processors and so entirely unlike conscious human organisms. Instead, we should ask how this new extension of human intelligence into digital technologies is altering *our* consciousness. How are we being transformed by our interaction with these machines? If we allow the hype to convince us that a computer could become conscious, the danger is that we inadvertently diminish human beings to the status of machines. In other words, our fascination with supposedly autonomous machines may lead us to gradually remake and degrade the richness and fluidity of our world and our relationships with one another in an effort to support the efficient functioning of our devices (Frank et al., 2024, p. 174).

This concern resonates with the integral philosopher William Irwin Thompson’s critique of AI. He argued (2003, p. 187):

“In order to grant consciousness to machines, the engineers first labor to subtract it from humans, as they work to foist upon philosophers a caricature of consciousness in the digital switches of weights and gates in neural nets. As the caricature goes into public circulation with the help of the media, it becomes an acceptable counterfeit currency, and the humanistic philosopher of mind soon finds himself replaced by the robotics scientist.”

Thompson's warning is not merely about the potential obsolescence of philosophers but about the broader societal implications of reducing human beings to the status of machines. This reduction not only undermines our understanding of what it means to be human but also risks eroding the very foundations of our ethical and moral intuitions, which are grounded in the recognition of the dignity and agency of human persons.

### **AI, Society, and the Future of Human Flourishing**

As we advance deeper into the age of AI, the question of how these technologies will impact human society becomes increasingly urgent. AI has the potential to bring about significant benefits, from improving healthcare outcomes through predictive diagnostics to enhancing education through personalized learning platforms. However, these benefits will only be realized if AI technologies are developed and implemented in ways that align with human values and promote human flourishing. For example, would it really be an improvement to have millions of children sitting at computer terminals running a simulation of a teacher, rather than exploring the wonders of the natural world with their full suite of human senses alongside human teachers capable of genuine empathy and shared wonder? What are the long-term consequences of replacing personal interaction between human beings with algorithmically driven screentime, particular in the case of children? Is learning just a matter of digesting new information, or is there also an important emotional and interpersonal dimension to developing a sound understanding of this wondrous world?

One of the most pressing concerns is the potential for AI to exacerbate existing social inequalities. As AI systems are increasingly used in areas such as criminal justice, hiring, and lending, there is a risk that these technologies will reinforce and even amplify biases that are already present in society. For example, if an AI system is trained on data that reflects historical

patterns of discrimination, it may inadvertently perpetuate those patterns in its decision-making processes. This is not just a technical issue but an ethical one, requiring careful consideration of the values that are embedded in AI systems and the societal impacts of their deployment.

Moreover, the integration of AI into the economy raises questions about the future of work and the distribution of wealth. As machines become capable of performing tasks that were once the exclusive domain of humans, there is a risk that large segments of the population will be displaced from their jobs, leading to increased economic inequality and social unrest. To address these challenges, we must think creatively about new economic models and social safety nets that can support individuals in a world where work as we currently understand it may no longer be the primary means of securing a livelihood. Again, not everything is an engineering problem that can be solved by adding more compute or by the next software update. AI raises existential and ethical challenges that only conscious human agents can resolve.

Finally, there is the question of how AI will affect our sense of identity and agency itself. As we increasingly rely on machines to act on our behalf, there is a risk that we will lose our sense of autonomy and self-determination. This is not just a theoretical concern but a practical one, as evidenced by the growing use of AI in areas such as surveillance, where individuals may find themselves subject to “decisions” made by algorithms that they do not understand and cannot challenge. I put “decisions” in scare quotes because the fact is that computers are incapable of consciously deciding anything. *We* are the ones who must decide what sort of power they are to have over our lives, and existing power inequalities obviously play a big part in shaping the technologies already being deployed by nation-states and big corporations.

## **Conclusion**

The race to create “Artificial General Intelligence” or conscious machines is more a deceptive advertising campaign than a technological challenge. AI and machine learning are powerful and exciting technologies that will continue to advance and find many valuable applications, transforming our economy in the coming years. The danger lies not just in the tools themselves but in our human relationship with them. By ensuring that AI technologies serve the interests of human beings and do not continue to be captured by corporate profit-seeking and military



applications, we can harness their potential to contribute to human flourishing rather than degrade it. Incorporating the insights of relevance realization further deepens our understanding of the limitations of AI and the unique qualities of human consciousness and agency. True relevance realization is an embodied, organismic process that cannot be fully captured by computational models.

As we move forward, it is essential to cultivate a philosophical perspective that recognizes the limits of AI and the unique qualities of human beings. This perspective must inform the development and deployment of AI technologies, guiding our society toward outcomes that enhance rather than diminish our collective well-being. By doing so, we can ensure that AI serves as a tool for human flourishing, rather than a force that undermines it.

## References

- Chalmers, D. (1996). *The conscious mind: In search of a fundamental theory*. Oxford University Press.
- Frank, A., Gleiser, M., & Thompson, E. (2024). *The blind spot: Why science cannot ignore human experience*. MIT Press.
- Jaeger, J., Riedl, A., Djedovic, A., Vervaeke, J., & Walsh, D. (2024). Naturalizing relevance realization: Why agency and cognition are fundamentally not computational. *Frontiers in Psychology*, 15. <https://doi.org/10.3389/fpsyg.2024.1362658>
- Kuhn, R. L. (2024). A landscape of consciousness: Toward a taxonomy of explanations and implications. *Progress in Biophysics and Molecular Biology*, 190, 28-169.
- Maturana, H., & Varela, F. (1980). *Autopoiesis and cognition: The realization of the living*. Reidel Publishing Co.
- Thompson, W. I. (2003). The Borg or Borges? *Journal of Consciousness Studies*, 10(4-5), 187.
- Whitehead, A. N. (1929). *Process and reality*. The Free Press.

Published in *AI Ethics in Practice: Navigating Academic Insight, Managerial Expertise, and Philosophical Inquiry* edited by Christian Hugo Hoffmann and Deepak Bansal (Spring 2025).

### **About the author**

Matthew David Segall, PhD is a transdisciplinary researcher, writer, teacher, and philosopher applying process-relational thought across the natural and social sciences, as well as to the study of consciousness. He is Associate Professor in the Philosophy, Cosmology, and Consciousness Department at California Institute of Integral Studies in San Francisco, and the Chair of the Science Advisory Committee for the Cobb Institute. Follow his work at [footnotes2plato.substack.com](https://footnotes2plato.substack.com)

**Name:** Matthew David Segall

**Position:** Associate Professor, Philosophy, Cosmology, and Consciousness Department at the California Institute of Integral Studies

**Address:** 1453 Mission St., San Francisco, CA 94103

**Email:** [msegall@ciis.edu](mailto:msegall@ciis.edu)

**Tel:** 415-906-1615