International Journal of Sustainability Management and Information Technologies
2025, Vol. 11, No. 2, pp. 67-75
https://doi.org/10.11648/j.ijsmit.20251102.11

Research Article

The Ghost in the Machine Is Us: Rhetoric, Reason, and the Search for a Humane AI

Mohammed Zeinu Hassen*

Department of Social Sciences, Addis Ababa Science and Technology University, Addis Ababa, Ethiopia

Abstract

The dominant paradigm of artificial intelligence (AI) has historically been grounded in a rationalist, computational model of thought that sees intelligence as a disembodied, logical process of symbol manipulation. This paper argues that this foundational philosophy, which severed logic from the social and contextual concerns of rhetoric, is the source of AI's most profound limitations and its most pressing ethical risks. Drawing on a historical analysis of AI's development and a philosophical critique informed by rhetoric and social constructionism, this article traces the trajectory of AI from its early, logic-based applications to its current role as a pervasive social force. It posits that the failures of AI in achieving common sense and the dangers it poses through algorithmic bias and threats to human rights are not mere technical flaws but direct consequences of its philosophical inheritance. The central thesis is that the "ghost in the machine" is not an emergent, independent consciousness, but rather a reflection of the human values, social structures, and rhetorical strategies we embed within its systems. Therefore, the search for a "humane AI" is not a technical problem to be solved but an ethical and philosophical commitment that requires re-integrating principles from the humanities to build systems that acknowledge the social, embodied, and narrative nature of true intelligence.

Keywords

Artificial Intelligence, Rhetoric, Social Constructionism, Ethics, Humane AI, Philosophy of Technology

1. Introduction

The modern idea of artificial intelligence is often of a dispassionate, logical mind. It is a thinking machine operating on pure reason. It is unburdened by the messy contingencies of human emotion and context. This image, popularized in science fiction and reinforced by the technological marvels of deep learning, presents AI as an entity on a linear path toward a superintelligence, a phenomenon with global implications [6, 7, 16]. This new intelligence will either save or supplant humanity. However, this perception of Al as a purely rationalist enterprise is a serious and dangerous misunderstanding. It is not an objective description but a historical and philosophical artifact. It is the result of a deliberate, centuries-long effort to sever logic from the human art of rhetoric. Classical AI was built upon a specific, and arguably flawed, model of the mind. It inherited a tradition that viewed intelligence as symbol manipulation, a formal process that could be abstracted from the body, from culture, and from the social world. As Lynette Hunter notes, this began with a historical shift that isolated a particular kind of analytical reasoning and placed it above all other forms of

*Corresponding author: mohammed.zeinu@aastu.edu.et (Mohammed Zeinu Hassen), mozhassen@gmail.com (Mohammed Zeinu Hassen)

Received: 19 August 2025; **Accepted:** 9 September 2025; **Published:** 9 October 2025

knowing [10]. This article will argue that this "original sin" of AI, its foundation in a decontextualized, demonstrative logic, is the root cause of its most notable failures and its most alarming ethical threats. From the brittleness of early expert systems to the insidious nature of modern algorithmic bias, the core problem remains the same. It is a failure to account for the fundamentally social, rhetorical, and embodied nature of human intelligence. This paper will trace the intellectual history of this flawed model, showing how early Al research reified the separation of logic from context. It will then introduce a counter-proposal, drawing on the work of philosophers and AI critics who advocate for a social constructionist view of intelligence. This alternative framework, articulated compellingly by William F. Clocksin, posits that intelligence is not an abstract capability of an individual mind [3]. It is instead constructed through the continual, ever-changing engagement with a social group. It is born of conversation, narrative, emotion, and persuasion. These are the very elements that the rationalist tradition sought to exclude. This paper will connect this philosophical debate to the urgent, real-world consequences we face today. The rise of AI in education, governance, and social life is not neutral; it is actively shaping our societies according to the values embedded in its code. The threats to human rights, the erosion of social cohesion, and the worsening of inequality are not bugs in the system. They are features of a model that instrumentalizes human experience. The central argument is this: the "ghost in the machine" is not some emergent, alien consciousness. The ghost is us. It is a reflection of our values, our biases, our social structures, and our rhetorical choices. Therefore, the search for a "humane Al" cannot be a purely technical job. It must be a philosophical and ethical one. It must be a project that reintegrates the wisdom of the humanities to build systems that reflect a more complete, compassionate, and just conception of what it means to be human. This analysis will proceed in three parts. It will first examine the "scientific rhetoric" of classical AI. It will show how the field defined itself through a commitment to formal logic and a behaviorist understanding of intelligence. Second, it will detail the turn toward context and social constructionism. This was a necessary shift prompted by the failures of the purely logical model. Third, it will demonstrate how the unresolved tensions of this philosophical debate manifest as the critical ethical and social challenges of contemporary AI. Ultimately, the paper concludes that a humane Al is only possible through an interdisciplinary approach that places rhetoric, ethics, and social theory at the heart of technological design.

2. The Scientific Rhetoric of Classical AI

The intellectual architecture of artificial intelligence was not built in a vacuum. It was the product of a specific historical and philosophical lineage. This lineage consciously separated analytical logic from the broader, more contextual world of human communication. This separation, as Lynette Hunter argues, was a key moment in Western thought that laid the groundwork for the scientific and, later, computational worldview [10].

2.1. The Severance of Logic from Rhetoric

For much of Western history, logic and rhetoric were intertwined. Persuasion required establishing grounds for an argument. A valid argument was essential for effective persuasion. However, a new kind of logic emerged, one that sought to purify itself from the ambiguities of human social life. This happened during the sixteenth and seventeenth centuries. Hunter describes this key shift: "During the sixteenth and seventeenth centuries, rhetoric in western Europe completely changed its face. At this time a logic isolated from rhetoric was proposed that was rational and analytical. This may have happened for social reasons because there was an emerging commercial class or group that perceived an education in rhetoric as an instrument of class constraint, which indeed it has often been. It has been suggested that the emerging group underwrote Peter Ramus's attempt to sever logic from rhetoric and then retain logic as the one valid component, generating a mode of pure reasoning" [10]. This new, "pure" logic was demonstrative and abstract. It moved from premise to conclusion in a reductive pathway. It created self-contained, tautological worlds where rules could be applied to achieve specific, predictable goals. Its grammar was exact. Its great promise was that it could proceed toward "truth" without the messy business of persuasion. As Hunter states, "Its rhetoric claims that there is no need for rhetoric. It takes its grounds for granted, builds self-defining tautologous worlds within which rules can be employed to achieve specific goals" [10]. This became the philosophical bedrock for modern science and, centuries later, for the emerging field of artificial intelligence. It is crucial to acknowledge that this rationalist approach was not without its merits; indeed, its profound success in certain domains is what made it so compelling. For well-defined, "closed-world" problems, such as mathematical proofs, logic puzzles, or games like chess, this demonstrative logic was incredibly powerful. It offered a clear and verifiable path to a correct solution, promising a form of intelligence that was predictable, efficient, and free from human ambiguity. This very success in structured environments cemented its status as the default paradigm for early AI researchers, making its subsequent limitations in the open, ambiguous world of human experience all the more significant.

2.2. The Turing Test and the Behaviorist Question

When Al emerged in the mid-20th century, it inherited this demonstrative, non-rhetorical model of thought. The central question was not about the nature of human under-

standing. It was about the performance of logical tasks. The famous "Turing Test," proposed by Alan Turing, epitomized this approach. The test did not ask what a machine was thinking or feeling. It simply asked whether its performance could be distinguished from that of a human. Hunter explains the two ways of interpreting this test: "The first is that of performance, where one asks the behaviorist question of 'does it act like a human being,' does it say the expected thing? The second and more interesting approach, which asks whether the underlying activity generating that performance has any parallels with human thought, is the cognitive question" [10]. Initially, Al focused almost exclusively on the first, behaviorist question. The goal was to build machines that could solve problems that humans usually solve. This led directly to the development of "expert systems," which were designed to mimic the logical processes of a human expert in a narrow domain.

2.3. The Symbol System Hypothesis: The Mind as a Computer

The guiding philosophy behind these expert systems was what Allen Newell and Herbert Simon would later call the "physical symbol system hypothesis." This was the explicit declaration of the classical AI model. It stated that intelligence, at its core, is simply symbol manipulation. It posits that a system can be intelligent if it can process physical symbols according to a set of formal rules. This applies to both a human brain and a digital computer. William F. Clocksin, in his analysis of this model, outlines its core components: "The two essential ingredients are representation, the idea that relevant aspects of the world need to be encoded in symbolic form in order to be usable, and the symbolic activity of inference, by which rational conclusions can be made by applying rules of inference to axioms" [3]. This hypothesis was revolutionary. It erased the distinction between mind and machine at a fundamental level. If thinking is just the formal manipulation of symbols, then a machine that manipulates symbols is, in fact, thinking. This worldview was deeply attractive. It made the vast, mysterious region of the human mind seem computationally tractable. It suggested that all of human reasoning, from medical diagnosis to legal argument, could eventually be reduced to a set of logical rules and algorithms. This confidence, however, masked a major philosophical assumption. It assumed that the messy, contextual world of human experience could be neatly encoded into symbolic form without losing its essential meaning. Many in the field were skeptical. Drew McDermott, a prominent AI researcher, famously addressed the overreach of this model. He argued that "the skimpy progress observed so far is no accident, and in fact it is going to be very difficult to do much better in the future. The reason is that the unspoken premise..., that a lot of reasoning can be analysed as deductive or approximately deductive, is erroneous".

2.4. Early AI in Practice: The Logic of History

The symbol system hypothesis controlled early Al research and application, despite its criticisms. The work of Richard Ennals in history teaching is a clear example of this model in action [5]. In the early 1980s, Ennals and his colleagues created educational programs. They used the logic-based programming language PROLOG. Their objective was to model historical reasoning as a computational process. They took historical sources like census data from 19th-century England. They translated this information into a database of logical statements. A census entry was represented in a formal, machine-readable structure, such as: (address) person (name relation condition age sex occupation birthplace). A specific entry could be: (48 Wordsley) person ((William Leek) Head M 56 Male Gardener (Stafford Wolverhampton)) [5]. The system could then use logical rules on this database to answer questions. This process simulated historical inquiry. Ennals proudly stated that in this system, "the starting-point is the human description: not the requirements of the computer system" [5]. However, the act of translating rich, ambiguous census data into a formal logic programming language is a major act of rhetorical and philosophical framing. The system can only reason within its predefined categories of age, sex, and occupation. It cannot account for unstated social realities. It cannot consider the cultural context or the human stories behind the data. Ennals even described using this model for a simulation of the Russian Revolution [5]. Here, political affiliations and historical events were coded as logical facts: Stalin member Bolsheviks; x wants revolution if x member Bolsheviks; Lenin tactics non-cooperation [5]. This example perfectly shows the classical AI model. It is an attempt to create an "intelligent" system with a self-contained, tautological world. The computer can make "inferences," such as that Stalin wants a revolution, because it is programmed with the rules that define that world. The system showed the power of formal logic, but it also revealed its serious limitations. It is a world without rhetoric, without ambiguity, and without the deep, contextual awareness that is the mark of true historical thinking. This was the intellectual inheritance of AI. It was a powerful but brittle foundation that the real world would soon challenge.

This diagram illustrates the paper's central thesis by tracing the intellectual lineage of Classical AI. It starts with the intertwined nature of logic and rhetoric in early thought, moves through the pivotal "Severance" of the two, and demonstrates how the path of "Pure Logic" directly formed the foundation for today's logic-based and decontextualized AI models.

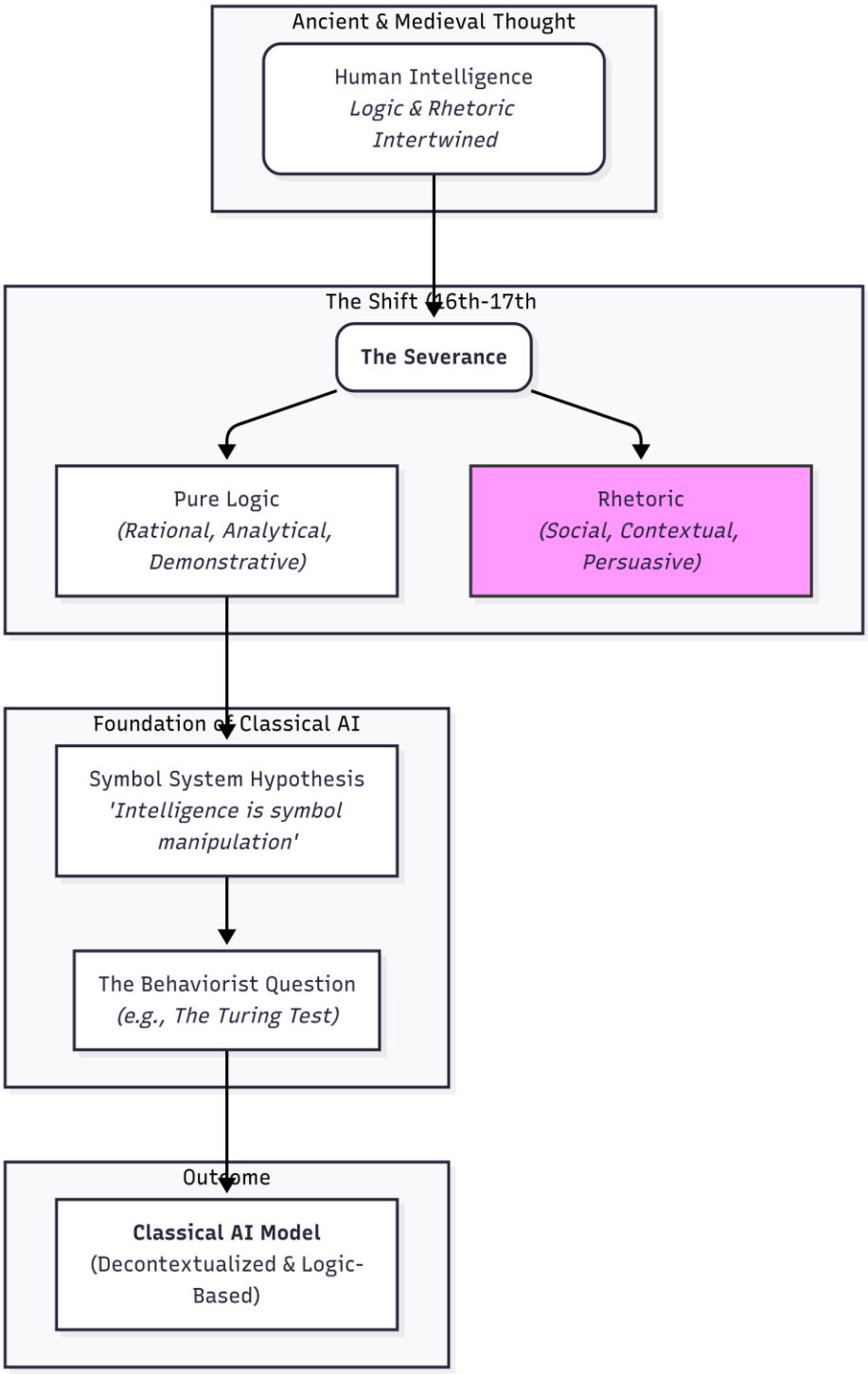

Figure 1. *Philosophical origins of Classical AI.*

3. The Turn to Context and the Human Element

The purely rationalist model of classical AI, with its focus on formal logic and symbol manipulation, had good outcomes in narrow, well-defined areas. It could play chess. It could diagnose specific diseases. It could also query structured databases. Yet, by the mid-1970s, it became clear this approach had reached a limit. The promise of general intelligence was out of reach. The reason was simple. The real world is not a closed, logical system. It is an open, ambiguous, and deeply contextual place. This fact caused a change in Al research. It was an unwilling move away from pure logic and toward the messy, human-centric ideas it had long tried to avoid.

3.1. The Crisis of Context and the "Commonsense" Problem

The greatest obstacle for classical AI was what researchers called the "commonsense problem." Machines that could do complex calculus were unable to understand simple truths about the physical and social world that a child takes for granted. For example, a system might know the chemical properties of water. It would not know that getting wet is unpleasant or that you cannot push a rope. This was the crisis that Hubert Dreyfus, a philosopher and firm critic of AI, had predicted. He argued that human intelligence is not based on manipulating abstract symbols but is situated and embodied, grounded in our physical and social experience of the world [3, 10]. This failure of the logical model led to a new focus on "natural language" and "knowledge representation." Al researchers started to admit that context mattered. Their initial approach, however, was still based on the old model. They tried to solve the context problem by creating more and bigger databases of facts. They turned context itself into another set of symbols to be manipulated. As Hunter notes, this was a turning point, but not a complete break: "In many ways this was the turning point away from Huizinga's theory of the ideological basis of games for large social structures to an interpretation of Wittgenstein's comments on games as individual strategies or designs. In the process there has been a general 'turn' to questions of 'natural' language" [10]. The attempt to formalize natural language and commonsense knowledge was monumentally difficult. It needed a different philosophical starting point. This new start could not see intelligence as an individual, abstract process, but as a fundamentally social event.

3.2. A New Proposal: Intelligence as Social Construction

The strongest alternative to the classical AI model comes from the field of social constructionism. This theory draws on the work of thinkers like Kenneth Gergen and Lev Vygotsky. It proposes that what we call "mind" or "intelligence" is not inside an individual's head. Instead, it is made and sustained through social interaction. William F. Clocksin, in his major paper "Artificial Intelligence and the Future," supports this view as the key to a more robust and humane AI [3]. He argues that the traditional AI focus on the individual, rational mind is a distraction. The real work of intelligence happens between people. Clocksin uses the work of Gergen to state this radical shift: "By contrast, constructionism describes the work of Gergen... It has also been called discursive psychology or social constructivism, which is probably the source of the confusion of terminology. The main concern of social constructionism are the processes by which human abilities, experiences, common sense and scientific knowledge are both produced in, and reproduce, human communities... The idea is that constructional processes are to be found in relationships, often discursive, between persons. These relationships embody situated reasoning; that is, a contextualized meaning, producing performance" [3]. From this viewpoint, intelligence is not something you have; it is something you do with other people. It is a performance. Meaning is not encoded in symbols but is negotiated in conversation. We understand the world not by using logical rules on internal representations, but by taking part in shared stories, conversations, and cultural practices. Clocksin quotes Gergen to show the radical decentering of the individual: "Instead, one's own role becomes that of a participant in social processes. The constructionist concept of the self and identity... assumes a 'transmission theory' of communication, in which messages are transmitted from one person to another... By contrast, the constructionist view sees communication as patterns of social action that are co-constructed in sequences of evocative and responsive acts; these patterns comprise an ecology that is our social world" [3]. This framework completely reverses the classical AI model. It suggests that to build an intelligent system, we should not be trying to model a single, rational mind. We should be trying to model the processes of social interaction through which minds are made. The fundamental unit of intelligence is not the symbol, but the relationship.

3.3. Rhetoric, Embodiment, and the Human Form of Life

This social constructionist view brings the concerns of rhetoric back to the front. If meaning is negotiated and not given, then persuasion, interpretation, and the establishment of common ground are not secondary activities. They are the very essence of intelligence. The goal is not to reach a single, objective "truth" but to engage in a continual process of making sense together. Furthermore, this social view of intelligence is tied to our physical, embodied existence. Our thinking is shaped by our needs, our desires, our emotions, and our physical interactions with the world. Clocksin points to the story of the humanoid robot Mr. Data from Star Trek: The Next Generation as a perfect illustration of the rationalist fallacy [3]. Data is a being of pure logic. Yet he is shown as fundamentally incomplete without his "emotion chip." The fiction intuits a philosophical truth: that intelligence without emotion, without embodiment, is not truly human. Dreyfus made this point central to his argument, stating that "intelligence must be situated, it cannot be separated from the rest of human life" [3]. A truly intelligent being needs a body. A body is not just a vessel for a computational brain. Our bodies are the source of the "know-how" that informs all our reasoning. This is the background of skills, cultural norms, and physical intuitions that we use without conscious thought. Clocksin quotes Dreyfus to make this point: "if one thinks of the importance of the sensory-motor skills in the development of our ability to recognize and cope with objects, or of the role of needs and desires in structuring all

social situations, or finally of the whole cultural background of human self-interpretation in our simply knowing how to pick out and use chairs, the idea that we can simply ignore this know-how while formalizing our intellectual understanding as a complex system of facts and rules is highly implausible" [3]. The turn to context, therefore, was not just a technical course correction for AI. It was a major philosophical challenge. It suggested that the project of creating an artificial "mind" could not succeed by focusing on logic alone. It required grappling with the social, rhetorical, and embodied realities of what it means to be intelligent. This challenge, however, has not been fully met. The legacy of the rationalist model continues. Its unresolved tensions are the source of the most critical ethical and social problems we face in the age of AI.

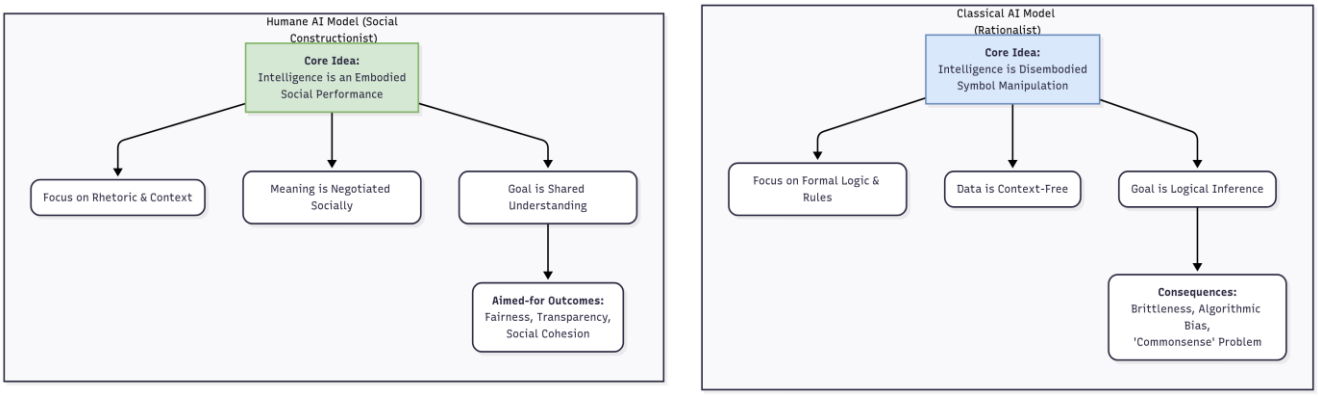

Figure 1. *Comparison of the two competing philosophies for Artificial Intelligence.*

This image contrasts the flawed Classical AI model, which treats intelligence as disembodied symbol manipulation, with the proposed Humane AI model, which understands intelligence as an embodied social performance. It shows how the classical model's focus on formal logic leads to problems like algorithmic bias, while the humane model's focus on rhetoric and context aims for positive outcomes like fairness and social cohesion.

4. Real-World Consequences of a Flawed Philosophy

While contemporary AI, driven by deep learning and neural networks, has moved beyond the explicit rule-based systems of classical AI, it has inherited its philosophical legacy. These modern systems still treat the world as a collection of context-free data points, encoding historical and social biases from their training data without understanding the rhetorical and social forces that produced them. Thus, the original sin of decontextualization persists, simply in a new, statistical form.

The philosophical debate between the classical, rationalist view of AI and the social constructionist alternative is not just an academic exercise. The legacy of the flawed rationalist model is directly responsible for the most urgent ethical and social crises of contemporary AI. We build systems based on the premise that intelligence is decontextualized symbol manipulation. This action creates technologies that are blind to the human contexts in which they operate. The consequences include algorithmic bias, the erosion of human rights, and the fragmentation of social cohesion. These are not accidental side effects. They are the logical outcome of a flawed design philosophy. The "ghost" in the machine is the haunting reflection of the human biases, power structures, and incomplete worldviews we have coded into its logic.

4.1. The Rationalization of Prejudice

Algorithmic bias is one of the most widely discussed dangers of modern AI. High-profile cases have shown that Al systems can continue and even worsen existing societal prejudices against women and people of color. These systems are used for loan applications and criminal justice. Facial recognition systems are known to be less accurate for women with darker skin. Predictive policing algorithms have been shown to disproportionately target minority neighborhoods. From the viewpoint of classical AI, this might appear as a data problem. It is a technical issue of "garbage in, garbage out." But from a humanist viewpoint, the problem is much deeper. It comes from the rationalist model's inability to account for context. The systems are not "biased" in the sense of having malice. They are simply carrying out their logical function on data that is a product of a biased world. Robert F. Murphy explains this in his review of Al in education: "The statistical models will encode the biases that are embedded in the training data. In general, the quality of prediction models built with machine-learning approaches is tied to the quality of the data that those models are trained on and the biases and competency of the humans who are evaluating and cor-

recting the models during the training phase. Biased decision-making algorithms can have serious implications for the people whose fate may depend on the output of these systems" [11]. The problem is that the AI system treats all data points as context-free symbols. It cannot understand the historical and social forces that made the data. For instance, an algorithm trained on historical hiring data might "learn" that men are more likely to be successful engineers. This is not because of any natural ability, but because of historical patterns of discrimination. The algorithm, operating on pure logic, simply finds a correlation and turns it into a rule. It rationalizes prejudice. Maria Stefania Cataleta points out this danger in the justice system: "'Gender shades' by the researcher Joy Buolamwini from MIT included) sustain that facial recognition can jeopardize our freedoms. Indeed, research conducted on different facial recognition systems... has shown that some ethnicities are treated in a more imprecise way compared to others. Notably, identification accuracy for Caucasian men was 99%, but only 34% for women with dark complexion. This is because algorithms of these systems are based on subject-data inputs which are prevalently male and of light complexion" [1]. This is a direct consequence of a model that separates logic from context. An attempt to build a system of "pure" reason has unintentionally built a system that is a perfect, uncritical mirror of our own imperfections.

4.2. The Fragility of Human Rights in the Face of the Algorithm

The rationalist model also poses a direct threat to the idea of human rights. The language of human rights is fundamentally rhetorical and contextual. It is based on ideas like dignity, autonomy, and justice. These ideas are negotiated in social and political discussion, not derived from logical axioms. Al systems, built on a foundation of demonstrative logic, have no natural capacity to understand these ideas. They operate in a world of efficiency, optimization, and prediction. In this world, human beings can easily become mere data points to be processed. Cataleta argues convincingly that this creates a fundamental asymmetry, a "fragility of human rights facing AI" [1]. She points to the use of Al in government surveillance and predictive justice as areas where this conflict is most sharp [9]. These systems, designed for the "presumed good" of security, can violate privacy, create new forms of discrimination, and weaken the presumption of innocence. The danger is that the algorithm's logic, which is efficient, scalable, and seemingly objective, can override the messier, more deliberative logic of human rights. Cataleta warns that this represents a major philosophical threat: "In the face of technological and scientific progress, the concept of a legal person is pushed to its limits, for 'the scientific and technological world, artificial in conceptual nature, come to encroach upon the already defined legal dimension of a person, an artificial concept in itself" [1]. This "encroachment" happens because the AI system does not view the human person as a subject with dignity, but as an object to be analyzed and predicted. The World Health Organization (WHO), in its 2021 report on the ethics of AI for health, echoes this concern [17]. The report notes the tension between Al principles and human rights standards. The report states that any ethical framework for AI must be grounded in "transparency, justice, fairness, non-maleficence and responsibility" [17]. These are not computational terms. They are rhetorical and ethical commitments that must be consciously designed into Al systems, often in direct opposition to the purely utilitarian logic that would otherwise guide them.

4.3. The Erosion of Social Cohesion and Digital Socialization

The application of AI in education and its impact on learning and socialization has been a topic of extensive research and development [2, 4, 12-15, 18]. The AI model's impact on education and socialization is perhaps its most subtle but far-reaching consequence. Peter D. Hershock argues in his analysis of the post-pandemic acceleration of intelligent technology that the shift toward digitally-mediated learning and social interaction carries major risks for social cohesion [8]. Human socialization has traditionally been an embodied, face-to-face process. It relies on the "mirror neuron system" to develop empathy, emotional fluency, and the capacity for social learning. Hershock warns that the move to online, AI-driven education could fundamentally change this process: "There is now a considerable body of neuroscientific evidence that asynchronous, digitally-mediated connectivity fails to activate the mirror neuron system that supports both verbal and non-verbal communication, thus compromising social learning, empathy, and emotional fluency... In short, the predominance of digital learning, rather than face-to-face learning, will have the effect of producing a social learning deficit" [8]. This is a direct consequence of building our educational and social technologies on a disembodied model of intelligence. If we believe that learning is simply the transmission of symbolic information, then an online platform seems like an efficient delivery mechanism. But if we take the social constructionist view, learning is an embodied, relational process. Then the limitations of these technologies become very clear. Hershock connects this to the broader market logic driving Big Tech's push into education [8]. This logic, focused on creating "personalized" experiences, risks isolating individuals. It replaces the shared, and sometimes difficult, work of social learning with a frictionless, consumer-oriented model. The result is a potential fragmentation of society. Individuals are socialized into niche digital communities rather than a broader, shared public sphere. This is the ultimate danger of the rationalist model. We design systems that treat humans as individual, disembodied minds. In doing so, we risk creating a world that re-

flects that very assumption. It is a world of isolated individuals, connected by logic but disconnected by the embodied, empathetic, and rhetorical bonds that create true community. The ghost in the machine is not just a reflection of our individual biases. It is a reflection of a flawed and incomplete philosophy of what it means to be human.

5. Conclusion

The journey of artificial intelligence, from its origins in the mid-20th century to its common presence today, has been a story of great technical achievement. But it has also been a story of major philosophical tension. The field was founded on a powerful but limited idea. This idea was that human intelligence could be replicated by separating logic from the difficulties of rhetoric, context, and embodiment. For decades, this rationalist model guided AI research. It produced systems that could excel at closed-world, logical tasks. Yet these systems remained brittle and uncomprehending when facing the open, ambiguous world of human experience. This paper has argued that the most pressing challenges of contemporary AI are not mere technical glitches to be patched. These challenges include algorithmic bias, the threat to human rights, and the erosion of social cohesion. They are the direct and predictable consequences of this founding philosophy. A system designed to see the world as a collection of context-free symbols will inevitably fail to understand the historical and social contexts that give those symbols meaning. A system that gives priority to logical efficiency will inevitably struggle to accommodate the non-negotiable principles of human dignity and justice. A system built on a disembodied model of mind will inevitably devalue the embodied, relational processes that are the foundation of human learning and community.

The central thesis of this paper has been that the "ghost in the machine" is, and always has been, us. The intelligence of our machines is a mirror. It reflects the values, assumptions, and rhetorical choices we make when we design them. If that reflection is often distorted and unsettling, it is because it is reflecting an incomplete and distorted view of ourselves. The classical AI model, in its search for pure reason, built a mirror that could only reflect the logical, analytical part of our humanity. It left the social, emotional, and embodied dimensions in shadow.

The search for a truly “humane AI,” therefore, cannot be a project for computer scientists and engineers alone. It requires a fundamental philosophical reframing. It needs a conscious effort to heal the historical split between logic and rhetoric. The alternative framework offered by social constructionism, as supported by critics like Clocksin, provides a powerful starting point. It reminds us that intelligence is not an abstract, individual possession. It is a social, relational, and embodied performance. It is created in conversation, sustained by narrative, and guided by the persuasive force of shared values. This means that the humanities are not a luxury in the age of AI; they are a necessity. The work of ethicists, rhetoricians, sociologists, and historians is essential to the project of building a better AI. We need their knowledge to understand the biases present in our data. We need it to state the human rights principles that must constrain our algorithms. We also need it to design technologies that build, rather than fragment, social cohesion. To achieve this, translating our philosophical commitment into practice requires moving beyond abstract goals and embedding humane principles directly into the AI development lifecycle. This involves several concrete actions:

Mandating Interdisciplinary Design: Ethicists, sociologists, and humanities experts must be integral members of development teams from the initial design phase, not just external reviewers of a finished product. Their role is to surface the embedded values and rhetorical choices in a system's architecture.

Prioritizing Context over Raw Data: Datasets should be treated not as collections of objective facts, but as social and historical artifacts. This means actively documenting and modeling the context, biases, and power structures that produced the data, making this context a primary feature for the AI to consider.

Expanding Metrics of Success: The performance of an AI system must be evaluated on more than just accuracy and efficiency. New, mandatory metrics must be developed to measure a system's fairness, transparency, contestability, and its tangible impact on human dignity and social cohesion.

Building for Deliberation, Not Just Decision: Systems, especially those in the public sphere, should be designed to augment human deliberation rather than replace it. They should be built to explain their reasoning in understandable terms and provide clear channels for users and communities to contest and correct algorithmic outputs.

As the World Health Organization report makes clear, the future of AI must be guided by a robust set of ethical principles. These are fairness, transparency, justice, and accountability. These principles cannot be an afterthought, a patch applied to a flawed system. They must be the foundation upon which the system is built. Ultimately, the challenge of AI is not about creating a machine that can think. It is about deciding what kind of thinking we value. Do we value the decontextualized, demonstrative logic of the classical model, with all its power and all its dangers? Or do we value the messy, contextual, and deeply human intelligence that is born of relationship and rhetoric? The future we build will depend on our answer.

Abbreviations

AI	Artificial Intelligence
MIT	Massachusetts Institute of Technology
WHO	World Health Organization

Author Contributions

Mohammed Zeinu Hassen is the sole author. The author read and approved the final manuscript.

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] Cataleta, M. S. (2020). Humane artificial intelligence: The fragility of human rights facing AI (Working Paper No. 02). East-West Center.

[2] Chai, C. S., Lin, P.-Y., Jong, M. S.-Y., Dai, Y., Chiu, T. K. F., Qin, J. (2021). Perceptions of and behavioral intentions towards learning artificial intelligence in primary school students. Educational Technology & Society, 24 (3), 89–101.

[3] Clocksin, W. F. (2003). Artificial intelligence and the future. Philosophical Transactions: Mathematical, Physical and Engineering Sciences, 361 (1809), 1721–1748.

[4] Deved˘zi´c, V. (2004). Web intelligence and artificial intelligence in education. Journal of Educational Technology & Society, 7 (4), 29–39.

[5] Ennals, R. (1982). History teaching and artificial intelligence. Teaching History, 33, 3–5.

[6] European Council on Foreign Relations. (2019). Harnessing artificial intelligence.

[7] Fricke, B. (2020). Artificial intelligence, 5G and the future balance of power. Konrad Adenauer Stiftung.

[8] Hershock, P. D. (2020). Humane artificial intelligence: Inequality, social cohesion and the post pandemic acceleration of intelligent technology (Working Paper No. 01). East-West Center.

[9] Hunter, A. P., Sheppard, L. R., Karl ́en, R., Hunter, A. P., Balieiro, L. (2018). Artificial intelligence and national security: The importance of the AI ecosystem. Center for Strategic and International Studies (CSIS).

[10] Hunter, L. (1991). Rhetoric and artificial intelligence. Rhetorica: A Journal of the History of Rhetoric, 9 (4), 317–340.

[11] Murphy, R. F. (2019). Artificial intelligence applications to support K–12 teachers and teaching: A review of promising applications, opportunities, and challenges. RAND Corporation.

[12] Perez, R. S., Seidel, R. J. (1990). Using artificial intelligence in education: Computer-based tools for instructional development. Educational Technology, 30 (3), 51–58.

[13] Roth, G. L. (1987). Artificial intelligence: Its future in technology education. The Journal of Epsilon Pi Tau, 13 (2), 47–50.

[14] Roth, G. L., McEwing, R. A. (1986). Artificial intelligence and vocational education: An impending confluence. Educational Horizons, 65 (1), 45–47.

[15] Scott, D. (2021). Technology, artificial intelligence and learning. In On learning: A general theory of objects and object-relations (pp. 181–193). UCL Press.

[16] Vempati, S. S. (2016). India and the artificial intelligence revolution. Carnegie Endowment for International Peace.

[17] World Health Organization. (2021). WHO consultation towards the development of guidance on ethics and governance of artificial intelligence for health.

[18] Yazdani, M., Lawler, R. W. (1986). Artificial intelligence and education: An overview. Instructional Science, 14 (3/4), 197–206.

Biography

Mohammed Zeinu Hassen is an Ethiopian philosopher and academic who earned both his Bachelor of Arts and Master of Arts degrees in philosophy from Addis Ababa University. He has taught at Aksum University and currently serves as a senior researcher at Addis Ababa Science and Technology University. Presently, he is pursuing a PhD in philosophy at the University of South Africa. His research interests encompass ethics, consciousness, human purpose, analytical philosophy, axiology, and the philosophy of science, with a strong emphasis on intercultural dialogue. Among his notable publications are "John Dewey's Philosophy of Education: A Critical Reflection" (2023), and "Cartesian Methodological Doubt Vis-à-Vis Pragmatism: An Approach to Epistemological Predicament" (2020).

Research Field

Mohammed Zeinu Hassen: Consciousness, Human purpose, Analytical philosophy, Axiology, Indigenous knowledge, Philosophy of science, Epistemology, Intercultural dialogue, Philosophy of education, AI and Public policy, Social and political philosophy.