ISSN:1738-6764, eISSN:2093-7504

Redefining AI Rationality

Yeorang Bak [1] and Jaewoong Kim [2,*]

[1] GSAIM, Chung-Ang University; Graduate student; youlight1111@gmail.com
[2] GSAIM, Chung-Ang University; Professor; kjw@cau.ac.kr
* Correspondence

https://doi.org/10.5392/IJoC.2025.21.2.088

Manuscript Received 9 September 2024; Received 25 March 2025; Accepted 21 April 2025

Abstract: *This paper recontextualized AI rationality through the lens of bounded rationality and media studies, arguing that AI's influence could extend well beyond algorithmic computations to reshape societal norms and communication processes. We examined how the transition from a purely utilitarian model of rationality to one constrained by data availability, computational limitations, and potential biases could impact both ethical considerations and practical outcomes. By framing AI as a mediator of human experience, our analysis revealed how inherent design choices could inadvertently perpetuate cultural or systemic biases. Drawing on philosophical perspectives and media theory, this study suggests that a more nuanced, context-aware model of bounded rationality can better guide AI systems toward responsible and inclusive decision-making. This paper concludes with recommendations for interdisciplinary collaborations—spanning ethics, engineering, and policy—to ensure that AI's development and deployment are aligned with human values, cultural diversity, and practical realities of real-world constraints.*

Keywords: AI; Rationality; Media; Decision-Making; AI Rationality

1. Introduction

Two principal methodologies are employed in the exploration of intelligence: the physical method and the evolutionary method. The physical method postulates that all reality is fundamentally composed of atoms, which has led to materialism gaining prominence in discussions of intelligence. According to this perspective, if mental states exist, they must inherently be reducible to physical states [1]. This perspective provides the foundation for the argument that understanding mental states in material terms could facilitate the creation of artificial intelligence that is equivalent to human intelligence. The evolutionary method, as an extension of materialism, postulates that human intelligence has evolved over approximately 5 billion years through a process of natural selection. This method further suggests that once such intelligence is understood, it could be harnessed to enhance evolutionary capabilities. The current debate on the reducibility of intelligence to physical states is dominated by these intertwined methods.

Despite these foundational approaches, no single method conclusively determines the nature of intelligence. Historical skepticism towards the nature of intelligence has precluded the formation of any universal presuppositions about it. John Searle referred to this absence as the "explanatory gap", while Ray Kurzweil emphasized that bridging this gap requires a leap of faith [2]. Because no definitive explanation for intelligence has been established, researchers must rely on a measure of faith when forming conclusions. Consequently, the investigation of intelligence becomes a philosophical endeavor, as science alone is insufficient to fully comprehend its intricacies. Kurzweil has designated the pursuit of understanding intelligence as "perhaps the most challenging and significant scientific endeavor of our era." Similarly, Stuart Russell has observed the pervasive uncertainty in our understanding, which has led to a cautious approach in discussions [3].

2. Extended Background

This absence of a shared framework has prompted the development of philosophical discourse, paving the way for further exploration as illustrated in Figure 1. Each perspective within the discourse on intelligence adheres to its own logic and rationale. Philosophical dichotomies such as dualism-monism and materialism-idealism frequently oppose each other and are subject to ongoing debate. It is noteworthy that a prevalent line of inquiry, rooted in the combined physics-evolutionary method, traces a genealogy from monism to materialism, physicalism (identity theory), token identity, functionalism, and ultimately computer functionalism. This continuum posits that the human mind and brain are equivalent, culminating in the theory of computer functionalism, which analogizes the brain to hardware and the mind to software. Nevertheless, such theories face significant opposition from the dualist viewpoint, which argues that computer functionalism fails to capture the qualitative aspects of human consciousness. For example, individual experiences of viewing the same painting differ substantially, challenging the functionalist view that reduces all such experiences to merely "seeing a painting." Those who oppose this approach, particularly those who reject functionalism, argue that it fails to take into account the complex nuances of sensory experiences.

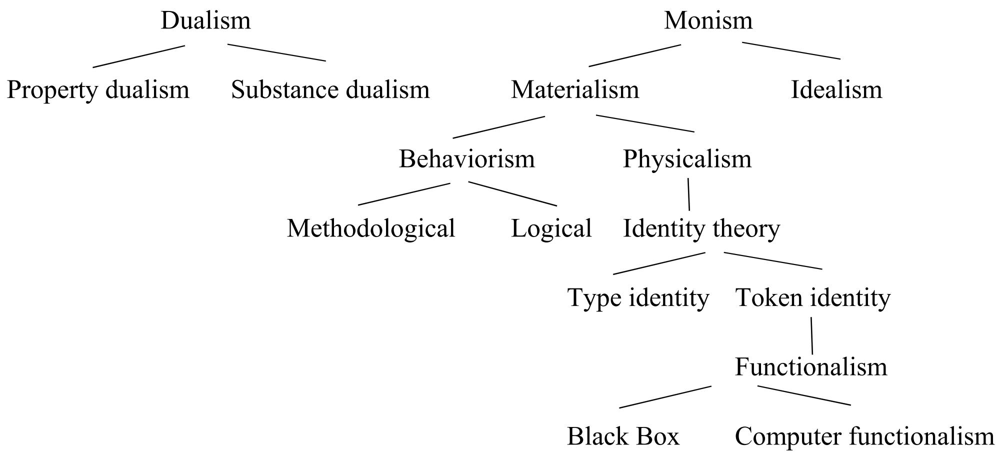

Figure 1. A philosophical genealogy of intelligence

The absence of a shared understanding of intelligence does not negate the necessity for establishing one. Similarly, the current inability to fully resolve the complexities of intelligence should not deter us from defining a trajectory for AI discourse. The question of what objectives should be set for future AI development remains crucial, especially in light of AI's existing commercial applications and its potential to address longstanding mysteries about human intelligence. Ultimately, the challenge of understanding intelligence represents an enduring quest that may eventually be unraveled or even redefined through discovery and innovation.

Consequently, researchers in the field of artificial intelligence have established a framework for categorizing the intellectual capabilities of AI, differentiating between human-like and rational functionalities. In examining the human-like aspects, researchers analyze and propose hypotheses based on observed human behaviors and cognitive processes. Conversely, the rational aspect incorporates a synthesis of mathematics, engineering, statistics, and economics. Psychology plays a significant role in the analysis of human-like traits, while statistics are more crucial to the analysis of rational processes. The essence of rationality in AI is encapsulated by the principle of "doing the right thing," a concept that varies widely. In particular, AI research adheres to the tenet of consequentialism [4], which assesses the rationality of an AI agent based on the outcomes it generates.

The epitome of rational thought and correct reasoning is exemplified by the syllogism. A paradigmatic illustration is the following: "All humans die; Socrates is a human; therefore, Socrates will die." This type of reasoning is fundamental to human thought and has facilitated the evolution of logic and the development of algorithms in artificial intelligence. However, a limitation of rational thinking is that it is based on certain assumptions that may not be entirely accurate. Statements such as "All humans die" are fundamentally probabilistic rather than absolute, and this level of certainty may not always be practical for resolving everyday challenges. Despite the capacity for rational thought that humans possess, inherent uncertainties frequently prevent such thought from consistently leading to rational behavior. In the context of AI, this uncertainty presents a particularly significant challenge, as AI's rationality is predominantly evaluated based on its outcomes, rather than the processes leading to those outcomes. Consequently, the rationality of AI is evaluated through a consequentialist lens, with the focus being on the results rather than the underlying behaviors.

> For each possible percept sequence, a rational agent should select an action that is expected to maximize its performance measure, given the evidence provided by the percept sequence and whatever built-in knowledge the agent has [4].

In essence, a rational AI is tasked with maximizing the information at its disposal in the face of uncertainty in order to achieve the best possible outcomes. This necessity transforms an uncertain environment into a testing ground where utility and probability replace traditional goals and logical reasoning as the cornerstones of AI rationality. By examining the evolution and application of utility and probability, we can gain a deeper understanding of and insight into the rational decision-making processes of AI systems. Consequently, the principles of utility and probability become pivotal in determining the optimal course of action in the face of uncertainty, serving as the foundational elements of rationality in AI systems.

In this context, AI rationality refers to an AI system's capacity to choose actions that maximize expected outcomes, while bounded rationality describes how real-world constraints (e.g., data limits, computational cost) can prevent perfect optimization.

3. Philosophical Foundations of AI Rationality

In the 18th century, Daniel Bernoulli introduced the concept of expected utility, which shifted the focus from monetary assessments of risk to their impact on individual utility. This foundational idea has influenced contemporary approaches to AI in risk assessment and decision-making, emphasizing outcomes that maximize individual utility over simple economic gains. For example, while a lottery ticket might offer a $20 prize with a 50% chance of winning, resulting in an expected monetary value of $10, Bernoulli argued that its true value to an individual depends on its utility. Consequently, for an individual experiencing severe financial distress, selling the ticket for $9 may offer a more immediate and tangible benefit than the potential but uncertain $20 prize. This perspective underscores that the value of risk is rooted in its expected utility, reflecting a more nuanced understanding of human decision-making that incorporates internal and subjective factors.

> The price of the item is dependent only on the thing itself and is equal for everyone; the utility, however, is dependent on the particular circumstances of the person making the estimate. Thus there is no doubt that a gain of one thousand ducats is more significant to a pauper than to a rich man though both gain the same amount [5].

Bernoulli's analysis revealed that the utility derived from a monetary increment does not scale linearly; instead, it tends to rise at a diminishing rate. This observation led to the understanding that utility, as a measure of value and satisfaction derived from goods or actions, is not directly observable and must instead be inferred from the preferences and choices that individuals make. This concept has profound implications for both economics and decision theory, as it suggests that utility is inherently subjective and varies between individuals.

The concept of utility has been a fundamental element in the evolution of rational decision-making, particularly in the context of uncertainty, a scenario that is frequently encountered in the field of AI research. In such contexts, the objective for AI systems is to maximize expected utility. This is achieved by combining the probabilities of various outcomes with the utility values assigned to those outcomes, thereby guiding the system toward the most beneficial actions. Nevertheless, the definition and measurement of utility remain complex. In practical terms, utility can be assessed through revealed preferences, which involve observing the choices made when multiple options are available. Nevertheless, the accurate quantification of utility in a manner that reflects subjective experience and satisfaction remains a challenging endeavor.

> Nature has placed mankind under the governance of two sovereign masters, pain and pleasure. It is for them alone to point out what we ought to do, as well as to determine what we shall do [6].

> The creed which accepts as the foundation of morals, Utility, or the Greatest Happiness Principle, holds that actions are right in proportion as they tend to promote happiness, wrong as they tent to produce the reverse of happiness [7].

Utilitarianism, which is based on the concept of utility, evaluates morality through the lens of maximizing individual interests and pleasures. This philosophical approach assumes that human actors are rational and employs a consequentialist framework, which determines the moral worth of an action based solely on its outcomes. John Stuart Mill, a prominent advocate for utilitarianism, highlighted the qualitative differences in utility and argued that rational individuals should aim to maximize their utility by preferring higher-quality pleasures over lesser ones. In accordance with the tenets of utilitarianism, the rational course of action is to select the pleasure that yields the greatest satisfaction or utility. It is crucial to note that the moral evaluation of an action under utilitarianism is not contingent on the character or intentions of the individual performing the action, but rather on the consequences of the action itself. This implies that an action is deemed morally right if it produces beneficial outcomes, regardless of whether the individual performing the action is perceived as morally good or bad. Consequently, consequentialism prioritizes the consequences of an action over its means, evaluating actions based on their capacity to achieve the most favorable overall results.

In the realms of rationality and consequentialism, the links between utilitarianism and AI agents are profound and intertwined. Ethical considerations are inextricably linked to the core functions of AI. As a rational agent, AI is tasked with maximizing utility, a principle that thrusts it into the center of ethical quandaries such as the trolley problem. This problem, a staple in discussions of ethics, highlights the utilitarian challenges AI faces in scenarios where it must choose whose utility to prioritize. Such scenarios raise questions about the inherent ethical responsibility of such systems. For example, when a self-driving car is compelled to choose between striking jaywalking pedestrians or endangering its passengers, or when an autonomous weapon must decide between targeting enemy soldiers at the risk of collateral damage to friendly forces, the AI's decision-making process becomes a direct reflection of utilitarian principles. These scenarios illustrate that while AI can optimize for expected utility, the determination of whose utility—be it an individual, a programmer, a nation, a corporation, or society at large—should be prioritized remains unresolved within the AI's programming. The outsourcing of ethical decisions to AI systems has the potential to absolve humans from direct responsibility, inadvertently diluting the collective moral conscience. Moreover, the reliance on AI to address ethical issues may result in the obfuscation of underlying ethical concerns, rather than their direct resolution. Consequently, one of the principal ethical issues with AI is its potential to "numb" our sense of ethics, thereby further exacerbating existing moral challenges rather than providing clear solutions. This situation highlights the necessity for a more sophisticated approach to AI development, one that incorporates ethical considerations into the design and deployment of these technologies.

Utilitarianism addresses ethical dilemmas by emphasizing the measurability of utility, a concept that necessitates the ability to compare and quantify the value of various outcomes. In order for utilitarianism to be effectively applied, it is essential to have a method for measuring utility in a way that allows for comparison across different scenarios. John Stuart Mill's approach to utility was groundbreaking; he contended that pleasure could be quantified and should be regarded as a direct consequence of utility rather than an abstract concept, transforming it into a tangible entity derived from utility calculations. Additionally, Jeremy Bentham made notable contributions to this field by developing a methodology for quantifying the subjective magnitude of utility. He proposed that the assessment of pleasure and pain should be guided by seven key factors: intensity, duration, certainty, propinquity, fecundity, purity, and extent. Bentham classified human experiences into 14 types of single pleasures and 12 types of single pains, noting that more complex experiences often arise from the combination of these basic types. Moreover, he identified 32 distinct sensitivities that could influence the intensity of pleasure and pain. A century later, the work of Mill and Bentham received statistical support, thereby confirming the veracity of their theories and providing empirical evidence that furthered the understanding of how utility could be quantified and applied in practical decision-making.

In the mid-20th century, John von Neumann and Oskar Morgenstern made a significant contribution to the field of utility theory by establishing its axiomatic foundations. In their foundational work, "Theory of Games and Economic Behavior," they introduced the concept of expected utility, which aimed to quantitatively assess and calculate utility in a systematic manner [8]. The objective was to provide a robust mathematical framework for understanding how individuals make rational decisions in the face of uncertainty. To this end, von Neumann and Morgenstern developed a set of axioms that describe the properties decision-makers are assumed to have when they evaluate choices. These axioms are of paramount importance to the expected utility theory and include: (1) Completeness: Every pair of outcomes can be compared, and one is either preferred to the other or regarded as equally desirable. (2) Transitivity: If an outcome A is preferred to outcome B and B is preferred to C, then A must be preferred to C. (3) Continuity: If an outcome A is preferred to B and B is preferred to C, there must be some mix of A and C that is equally preferable to B. (4) Independence: The preference between any two prospects should remain unchanged if they are combined with a third prospect in the same proportion. These axioms serve as the foundation of the expected utility function, which von Neumann and Morgenstern utilized to predict the behavior of rational agents based on the maximization of their expected utility. The establishment of this axiomatic system has had a profound impact on economics, psychology, and decision theory. It provides a clear and calculable method to gauge rational choice [9]. Furthermore, this foundation has also influenced modern approaches to risk assessment, economic modeling, and strategic decision-making in various fields.

The expected utility theory, as formulated by Neumann and Morgenstern, postulates that rational agents whose preferences align with a specific set of axioms will invariably act in ways that maximize their expected utility. This principle is fundamental to utility theory and serves to illustrate the predictability of decision-making processes among rational individuals. By defining a computable function—the expected utility function—they provided a powerful mathematical tool that quantifies how choices are made when outcomes are uncertain and varied. This computable nature of the expected utility function is of pivotal importance because it translates qualitative preferences into quantifiable data that can be analyzed and manipulated mathematically. This function has subsequently facilitated the development of Bayesian functions, which are crucial in contexts of uncertainty. Bayesian approaches to decision-making refine the expected utility framework by incorporating probability distributions to model uncertainty in a more dynamic manner. This is particularly pertinent in contexts where outcomes are not deterministic but probabilistic. Bayesian functions, which embody the principles of rational choice under uncertainty, have become instrumental in various applications within artificial intelligence. These functions permit AI systems to make decisions that optimally balance expected benefits against risks, accounting for the likelihood of different outcomes. This approach is fundamental to enhancing the intelligence and decision-making capabilities of AI systems across a broad spectrum of scenarios. It is applicable to a range of domains, including autonomous vehicles navigating unpredictable environments and financial systems managing investment risks [10].

4. Rethinking AI Rationality

The concept of calculating utility is deeply embedded in the philosophy of utilitarianism, which posits that such calculations can serve as an optimal tool for decision-making among rational actors. Nevertheless, Daniel Dennett identifies a significant challenge to this assumption, noting that "the ethically relevant effects of our contemplated actions are bound to be incalculable unless we place arbitrary limits on them [11]". This assertion underscores the inherent limitations of attempting to quantify all variables that influence the outcomes of decisions. Consider the hypothetical scenario of a ceasefire between North and South Korea. The assessment of its benefits must consider more than a simple calculation of immediate economic gains. A more nuanced approach is required to consider the long-term social impacts, regional stability, and moral implications of such a decision. These factors demonstrate the intricacy of utilitarian calculations, which frequently fail to encompass the full spectrum of human values and ethics in real-world scenarios.

This complexity gives rise to the concept of bounded rationality, which was first proposed by Herbert Simon and subsequently developed by researchers such as Reinhard Selten and Bryan D. Jones. Bounded rationality acknowledges that while humans strive to make rational decisions, they do so within the constraints of their cognitive and emotional capacities. In contrast to the ideal of perfect rationality proposed by classical utilitarianism, bounded rationality acknowledges that decision-makers are "intentionally rational" but constrained. They are goal-oriented and adaptive, yet they may still fail to make optimal decisions due to their inherent limitations [12, 13]. In the field of artificial intelligence, there is a growing tendency to move away from the classical view of utilitarian rationality and towards one that incorporates principles of bounded rationality. This transition reflects an acknowledgment that AI systems, much like humans, are constrained by various factors and operate in an uncertain environment.

In this respect, limited rationality in AI explicitly recognizes how an AI agent's constraints—such as incomplete data or restricted computation—shape its ethical and practical decisions, mirroring human bounded rationality. The development of AI models is being pursued with the objective of enabling them to make decisions that are not only rational in a classical sense but also adaptive and responsive to the complexities of real-world environments. This approach aims to create AI systems that more accurately mimic human decision-making processes, recognizing the bounds of their capabilities and the intricate contexts in which they operate.

Recent developments in AI research have focused on the optimization of system performance when resources are constrained. This concept is explored in "Computation-Limited Bayesian Updating" by Zhu et al [14]. This study examines the efficacy with which AI systems can update their beliefs or knowledge bases in light of new information when confronted with constraints on data availability and computational resources. Just as humans must adjust their understanding of the world based on incoming data, AI must also do so, but this adaptive process is often hindered by the finite nature of its processing capabilities. Similarly, "Bayesian Reinforcement Learning with Limited Cognitive Load" by Arumugam et al [15]. proposes a framework tailored for scenarios where resources, such as computational power and data, are sparse. This framework adapts the traditional Bayesian approach, which typically presumes a level of rationality and resource availability that may not be realistic in practical applications. In contrast, it presents a constrained Bayesian reinforcement learning model that addresses the constraints similar to those encountered by biological systems, including humans. This model demonstrates how AI can still learn and make decisions effectively, albeit within the limits of the available resources. These approaches represent a shift in AI research towards more realistic models of decision-making and learning, acknowledging the constraints under which both biological and artificial systems operate.

By incorporating these limitations into the foundational models of AI learning and decision-making, researchers aim to develop systems that are not only more efficient under resource constraints but also potentially more aligned with human-like processing capabilities. The evolving concept of AI rationality now places greater emphasis on optimizing decisions within resource constraints, rather than solely on maximizing expected utility. This aligns more closely with the principles of economic optimization and bounded rationality. This shift acknowledges that AI systems must operate efficiently within the finite limits of available resources, mirroring the decision-making constraints that humans face.

However, this reorientation towards bounded rationality introduces significant challenges, particularly in the form of bias. As observed by Nishant et al [16], when AI algorithms operate with limited or non-diverse data sets, they are prone to developing a narrow perspective that fails to represent the broader array of possible scenarios. This limitation necessitates that AI systems rely excessively on established patterns and rules, which can grossly oversimplify complex human situations, thus leading to biased outcomes. Such biases are not merely statistical anomalies but rather systemic issues that can perpetuate inequities and flawed decision-making. Furthermore, the constrained data environment typical of bounded rationality may result in AI systems failing to fully comprehend or incorporate all aspects of a given decision, which can lead to decisions that lack a holistic understanding. These shortcomings can serve to amplify specific biases or errors, particularly in situations where the cognitive demands exceed the AI's processing capabilities. This underscores the inherent risk in the current development of AI systems. While striving for efficiency, they must also navigate the delicate balance between operational constraints and the need for fairness and comprehensive understanding.

Indeed, the transition from an idealized form of full rationality to a more realistic bounded rationality presents significant advantages. This shift acknowledges the inherent limitations shared by both AI systems and humans in processing information, which is critical for creating more realistic and effective AI models. Adopting bounded rationality in AI development facilitates a more holistic and practical approach to system design, aligning with the intrinsic limitations of human decision-making processes, which are inevitably influenced by constraints in knowledge, resources, and cognitive capabilities. This perspective enables the development of AI systems that are better attuned to the nuances and complexities of real-world situations, where data is often incomplete and computational resources are finite.

The redefinition of AI rationality also challenges the traditional dichotomy between rationality and humanity. This suggests that AI, like humans, operates under constraints and imperfections. The recognition that both AI and humans make decisions within a framework of bounded rationality helps to bridge the conceptual gap between human cognitive processes and artificial intelligence. This alignment is of paramount importance for the advancement of AI systems that are capable of more effectively mimicking human-like decision-making and adapting to similar constraints. Furthermore, this approach emphasises the shared challenges of decision-making within limits, demonstrating that both AI and humans operate under similar constraints but may employ different strategies to optimise decision outcomes. The recognition of these parallels can facilitate the integration of AI systems into human-centric environments, thereby rendering AI a more useful and adaptable tool.

Ultimately, the incorporation of bounded rationality within AI can facilitate the generation of innovations that respect the limitations of both artificial and human intelligence. This approach can also foster a more profound comprehension of the inherent boundaries and potentialities of both. This understanding is of paramount importance for navigating the ethical, practical, and cognitive challenges that arise as AI systems become increasingly prevalent in society. The insights gained from recognizing and embracing these limits can significantly enhance the design and deployment of AI, ensuring that it contributes positively to the solution of complex problems in ways that are informed by both human and machine intelligence.

5. AI rationality as a mediator

An examination of AI rationality through the lens of "media" offers a novel perspective on how artificial intelligence interprets, processes, and communicates information. This perspective broadens the discussion from a simple comparison of human and AI intelligence to an exploration of the ways in which AI as a medium affects and transforms the communication and perception of information. When AI is regarded as a medium, its role extends beyond mere calculation or decision-making; it becomes a filter through which information is presented and understood. The mediation of AI impacts the "rational nature" of the data it processes, thereby inherently altering the content's original meaning or intent due to the AI's programmed biases, design constraints, and operational frameworks. In this context, the concept of limited rationality refers to the AI's capacity to process and interpret information within the constraints of its design and programming. An understanding of AI as a medium thus presents specific challenges and solutions related to AI rationality. This prompts us to examine the manner in which the design of AI affects the manner in which information is curated, prioritized, and presented. This raises critical concerns about the ethical implications of AI-mediated communication, particularly regarding transparency, accountability, and the potential for unintentional biases that could mislead or misinform users. By focusing on the role of AI as a mediator of information, we gain valuable insights into the ways in which artificial intelligence reshapes communication dynamics and the perception of reality itself.

> How come there are minds? And how is it possible for minds to ask and answer this question? The short answer is that minds evolved and created thinking tools that eventually enabled minds to know how minds evolved, and even to know how these tools enabled them to know what minds are. What thinking tools? The simplest, on which all the others depend in various ways, are spoken words [11].

> Of course, we are able to perform this task easily if we write down the sequence and then read it backward. In doing so we are using a technology-written language-to compensate for one of the limitations of our unaided thinking, albeit a very early tool. (It was our second invention, with spoken language as the first.) This is why we invent tools-to compensate for our shortcomings [2].

Ray Kurzweil and Daniel Dennett discuss the role of tools in enhancing cognitive functions. However, their perspective might be seen as somewhat reductive, particularly when considering the broader implications of media as transformative agents in society. While they acknowledge the utility of tools like language, writing, microscopes, cameras, and computers in extending the capabilities of the human brain, their analysis primarily frames these tools—including AI—as functional aids that compensate for human limitations rather than as transformative elements that could fundamentally reshape human experience and societal structures. In this context, Dennett and Kurzweil's approach could be interpreted as overlooking the profound impact that media, conceived broadly to include AI, have on cultural evolution and global communication networks. In its fullest sense, media not only enhances individual cognitive capabilities but also redefines how humans interact, communicate, and evolve culturally. It functions as a catalyst for cultural shifts and new forms of interaction, rather than merely serving as an adjunct to human intellect.

It is therefore imperative to recognize that this mediation by AI is fundamentally distinct from human cognitive processes. It is important to note that AI does not merely replicate human thought; rather, it processes and presents information within the constraints of its programming and operational capacities.

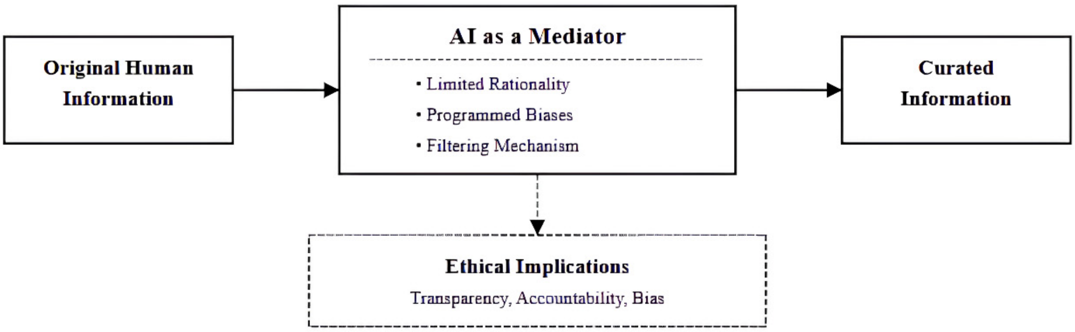

Figure 2. Conceptual Structure of AI Mediation

It functions as a resource-conscious optimization machine, perpetually striving to achieve a balance between computational efficiency and the necessity of delivering meaningful outcomes. Nevertheless, the role of AI as a mediator also introduces critical ethical considerations. As AI systems are designed with predetermined algorithms and data sets, they are susceptible to incorporating the inherent biases present in their design and the data they are fed. This renders AI a biased actor in the mediation process, one that can potentially perpetuate existing prejudices and misrepresentations unless these systems are carefully designed with awareness of and safeguards against such biases. The recognition of AI as a system with bounded rationality compels us to consider the broader implications of its mediation role. This prompts a critical examination of the ways in which AI influences our perception of the world and the potential impact on societal norms and individual behaviors. Consequently, the study of AI in media studies not only enhances our understanding of its technical capabilities but also its cultural, ethical, and social impacts. This comprehensive approach can facilitate the alignment of AI development with broader human values and societal goals, thereby promoting a balanced integration of this powerful technology into the fabric of daily life.

The integration of the concept of AI as media into media studies not only alters our comprehension of its functional role but also intensifies our engagement with philosophical and existential inquiries concerning its impact on human existence. By viewing AI through this lens, we can consider it not only as a tool or autonomous agent, but also as a mediator of human experience and existence itself. Adopting the view put forth by Flusser & Novaes [17], which posits that media serves as an anti-entropic force against the annihilation of existence, leads to the conclusion that AI, as a form of media, participates in shaping, extending, and sometimes complicating our existence. It does so in ways that are intelligent, boundedly rational, economically optimized, and inevitably biased. Each of these attributes contributes to the distinctive manner in which AI mediates human life, resulting in a spectrum of impacts. These include enhancing efficiency and decision-making capabilities, perpetuating biases, and creating new forms of dependence.

The advent of AI has prompted a series of profound inquiries regarding the nature of existence and the role of AI in shaping our understanding of the self and the world. The influence of AI on our perceptions of self and the world is a crucial area of enquiry. Furthermore, it is necessary to consider the impact of AI on decision-making processes, cultural norms, and interpersonal interactions. Moreover, it is crucial to examine how AI affects our responses to life's fundamental challenges and opportunities. The discussion must therefore extend beyond the mere consideration of the biases and ethical concerns associated with AI algorithms to encompass the potential of AI as a medium for mediating human existence. What are the desired outcomes of this mediation? It is desirable that AI should enhance human capabilities and experiences without overriding human autonomy or exacerbating social inequalities. It should facilitate the creation of environments that facilitate a more profound understanding and interaction with the world.

6. Conclusion

This paper has underscored the importance of moving beyond a purely utilitarian or consequentialist view of AI rationality and instead embracing a bounded rationality framework that recognizes real-world constraints. By conceptualizing AI as a mediator, we have illustrated how its role transcends algorithmic efficiency to reshape communication practices and societal norms. Such a perspective highlights that biases may stem not only from technical deficiencies—such as inadequate training data or constrained computational resources—but also from design choices and cultural assumptions embedded in AI systems.

Notwithstanding these insights, the study is primarily theoretical, relying on philosophical and conceptual discussions rather than empirical validation. This scope limits our capacity to demonstrate how bounded rationality operates in actual AI deployments across diverse domains like healthcare, finance, or social media. Another limitation is our focus on general AI architectures, which may not fully capture sector-specific complexities and regulatory environments.

Future research should therefore adopt a multidisciplinary strategy, engaging engineers, ethicists, media theorists, and policymakers. Empirical case studies are crucial for examining how AI's bounded rationality manifests in practice, particularly regarding ethical safeguards, bias mitigation, and transparent governance. Moreover, culturally nuanced investigations could clarify how AI design intersects with local values and social realities. By integrating media studies with technical and philosophical insights, we can cultivate AI systems that not only perform effectively under real-world constraints but also respect human dignity, promote equitable outcomes, and enhance collective well-being.

Acknowledgments: All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

[1] J. R. Searle, "Minds, Brains, and Programs," Behavioral and Brain Sciences, vol. 3, no. 3, pp. 417-424, 1980, doi: https://doi.org/10.1017/S0140525X00005756.
[2] R. Kurzweil, How to Create a Mind: The Secret of Human Thought Revealed, Penguin, 2013.
[3] S. Russell, Human Compatible: AI and the Problem of Control, Penguin UK, 2019.
[4] S. J. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, Pearson, 2016.
[5] D. Bernoulli, Exposition of a New Theory on the Measurement of Risk, in The Kelly Capital Growth Investment Criterion: Theory and Practice, World Scientific, pp. 11-24, 2011.
[6] J. Bentham, The Collected Works of Jeremy Bentham: An Introduction to the Principles of Morals and Legislation, Clarendon Press, 1996.
[7] J. S. Mill, Utilitarianism, in Seven Masterpieces of Philosophy, Routledge, pp. 329-375, 2016.
[8] J. Von Neumann and O. Morgenstern, Theory of Games and Economic Behavior: 60th Anniversary Commemorative Edition, Princeton University Press, 2007.
[9] S. Ferrari-Toniolo, P. M. Bujold, F. Grabenhorst, R. Báez-Mendoza, and W. Schultz, "Nonhuman primates satisfy utility maximization in compliance with the continuity axiom of expected utility theory," Journal of Neuroscience, vol. 41, no. 13, pp. 2964-2979, 2021, doi: https://doi.org/10.1523/JNEUROSCI.0955-20.2020.
[10] O. Pourret, P. Na, and B. Marcot, Bayesian Networks: A Practical Guide to Applications, John Wiley & Sons, 2008.
[11] D. Dennett, Moral First Aid Manual, Tanner Lectures on Human Values, III, Salt Lake City, 1988.
[12] B. D. Jones, "Bounded rationality," Annual Review of Political Science, vol. 2, no. 1, pp. 297-321, 1999, doi: https://doi.org/10.1146/annurev.polisci.2.1.297.
[13] R. Selten, "Bounded rationality," Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft, vol. 146, no. 4, pp. 649-658, 1990.
[14] J. Q. Zhu, A. Sanborn, N. Chater, and T. Griffiths, "Computation-Limited Bayesian Updating," Proceedings of the Annual Meeting of the Cognitive Science Society, 2023.
[15] D. Arumugam, M. K. Ho, N. D. Goodman, and B. Van Roy, "Bayesian reinforcement learning with limited cognitive load," Open Mind, vol. 8, pp. 395-438, 2024.
[16] R. Nishant, D. Schneckenberg, and M. Ravishankar, "The formal rationality of artificial intelligence-based algorithms and the problem of bias," Journal of Information Technology, vol. 39, no. 1, pp. 19-40, 2024.
[17] V. Flusser and R. M. Novaes, Communicology: Mutations in Human Relations?, Stanford University Press, 2022.