ABSTRACT

Artificial Intelligence (AI) originally served as a tool for cognitive science, providing computational models and frameworks, but now the path has changed and the goal now is to create systems that mimic human cognition. The authors argue that true human-like AI is unfeasible, current approaches misrepresent cognition and AI should (re)focus as a conceptual tool.

INTRODUCTION

The term ‘Artificial Intelligence’ has various meanings, from recreating intelligence in artificial systems to referring to the research fields pursuing such systems. Domain-specific AI performs narrowly defined tasks, like playing chess, while Artificial General Intelligence (AGI) aspires to human level cognition – a goal that remains computationally challenging. Early AI research was connected to cognitive science, using computational models to understand human thinking, but this focus has shifted as AI increasingly relies on ML, weakening ties to psychology. AI has also faced criticism for promoting harmful ideologies and prioritizing engineering goals over its interdisciplinary origins. The current AI hype, created by big tech companies and focused on technological advancements and profit, has sidelined AI as a theoretical tool for cognitive science. As AI systems are increasingly seen as instantiating parts of minds, this creates distorted views of cognition, weakening our theoretical understanding.

This paper seeks to address this issue in 2 steps. First, the authors challenge the dominant view that achieving human-like cognition in AI is a short-term inevitability, presenting a mathematical proof of the intractability of this task. AI systems cannot , for now, scale to human-level cognition. Second, they propose the return of the AI theoretical tool, emphasizing the importance of understanding cognition as computation without mistaking this for the practical replication of human minds in machines.

OVERVIEW

The paper is structured in 2 acts. In ACT 1, the authors present a thought experiment showing that the pursuit of AGI is theoretically intractable, using a formal proof to demonstrate the computational limits of current AI approaches. In ACT 2, they discuss the original role of AI in cognitive science, acknowledging its potential while addressing past mistakes, and propose ways to reclaim AI as a theoretical tool for understanding cognition.

ACT 1

ACT 1 begins by addressing the dominant AI narrative that claims AGI is inevitable and just around the corner due to rapid advancements in ML. The paper challenges this by introducing a thought experiment featuring a fictive AI engineer, Dr. Ingenia, who attempts to build AGI under idealized conditions. The authors then argue that while AGI may seem possible in principle, it is computationally intractable in practice, using formal proofs to demonstrate the inherent difficulty in realizing human-level AI.

FORMALISING AI by Engineering

This section formalizes the problem Dr. Ingenia faces in building AI by learning. Dr. Ingenia’s task is to find an algorithm that approximates human-like behaviors under ideal conditions, such as perfect data. The constraints placed on the problem involve ensuring the algorithm is expressible within a fixed size and tractable in terms of time complexity. The formalized AI-by-Learning problem includes a probabilistic requirement for the algorithm to perform human-like behaviors at a level significantly better than chance. The goal is to explore the computational challenges of achieving this with machine learning approaches.

INGENIA THEOREM

The Ingenia Theorem demonstrates that AI-by-Learning is computationally intractable. It involves the theoretical problem called Perfect-vs-Chance, which determines whether an algorithm can perfectly predict a distribution or whether the predicted accuracy cannot exceed chance. The proof of the Ingenia Theorem shows that if AI-by-Learning were tractable, it would imply the tractability of Perfect-vs-Chance, leading to a logical contradiction. Since Perfect-vs-Chance is intractable, AI-by-Learning must also be intractable.

IMPLICATIONS

The Ingenia Theorem emphasiezes the intractability of creating human-like AI through learning. While AI systems can seem impressive in specific contexts, scaling them to a human-level complexity is computationally infeasible. The theorem challenges claims that human-like AI is inevitable, as any tractable procedure must guarantee success with realistic constraints, which the theorem shows is impossible. Claims about AI approximating human cognition or being replacements for humans in tasks or experiments often distort our understanding of human cognitive capacities. This misrepresentation weakens theretical approaches in cognitive science. Additionally, the divide-and-conquer approach to modeling human cognition is also flawed, as assembling tractable sub-models contradicts the proven intractability of modeling cognition as a whole.

ACT 2

ACT 2 critiques makeism, the belief that building cognition in machines is sufficient for understanding it. The authors argue that this approach is flawed because cognition cannot feasibly be recreated computationally, as demonstrated by the Ingenia Theorem in Act 1. Despite the appeal of simulating human-like cognition, makeism’s failure led to its decline among early cognitive scientists and may mislead modern AI efforts.

The authors propose reclaiming AI’s role in cognitive science by focusing on computationalism, which uses AI concepts to develop and evaluate theories rather than attempting to build cognition itself (makeism). Computationalism remains valuable for exploring cognitive processes beyond the limits of experimental psychology and for learning from theoretical failures. A non-makeist approach treats AI as a framework for distinguishing between algorithms, the capacities they enable, and their physical implementations, allowing theoretical progress without conflating simulation with explanation.

Capacities as Problems

Computational-level modelling in cognitive science frames cognitive capacities as computational problems, allowing precise formal theoretical development without depending on algorithmic or implementation-level assumptions. This approach supports evaluating claims about cognition, such as explanatory adequacy, tractability, and evolvability, using mathematical proofs and formal models.

Algorithms and Simulations

Algorithmic and implementation level models are critically assessed through computational tools, often aided by computer simulations, These simulations serve as theoretical tools for proof of concept, assessing explanatory adequacy or exploring mechanistic possibilities. Unlike makeism, this approach avoids conflating models with reality, treating simulations as abstract, idealized extensions of human reasoning to explore what-if scenarios.

Underdetermination

Cognitive capacities are multiply realizable, meaning the same problem can be solved by different algorithms and physical systems. This leads to massive underdetermination: observed behaviors consistend with a computational theory cannot definitively reveal the underlying proceses. While computational theories remain conjectural due to their alignment with infinite potential models, tools like computational frameworks help evaluate and constrain plausible theories.

Computational Realisability

Computational level theories must be constrained by the requirements of computability (in principle realizability) and tractability (in practice realizability). While such theories often postulate uncompuable or intractable problems, this provides an opportunity to refine them through minimal revisions, preserving core intuitions while ensuring computational feasibility. This process can yield new insights, predictions or research directions. However, it cannot fully eliminate the inherent underdetermination of theory by data, highlighting the limits of computational modeling in understanding cognition.

Slow (Computational Cognitive) Science

Progress in understanding and explaining minds requires a deliberate, slow approach, focused on careful theoretical development and computational modeling. Unlike the rapid advancements in ML, this method builds robust, meaningful knowledge by addressing limitations and consequences, avoiding shortcuts or confusion between ML and theory. This deliberate pace ensures AI serves as a valuable theoretical tool in cognitive science.

CONCLUSION

AI cannot feasibly replicate human cognition due to computational intractability, even under ideal conditons. Treating AI systems as substitutes for human cognition misleads science and distorts understanding. AI should refocus as a theoretical tool for studying cognition rather than attempting to recreate it.