The Epistemological Asymmetry of Artificial Intelligence across Science and Mathematics

Himanshu Singh *

*ORCID: 0000-0001-8816-1340
*Website: https://himanshuvnm.github.io/

ABSTRACT

Artificial Intelligence (AI) architectures are increasingly capable of contributing to scientific discovery across a broad range of experimental disciplines, particularly in physics, chemistry, biology and engineering. In these domains, machine learning models can easily assist with hypothesis generation, detailed pattern discovery and complex design spaces exploration. In strict contrast to this, comparable advances in mathematics remain somewhat limited, despite significant advancement in automated theorem proving and formal verification. This perspective argues that the contrast reflects a deeper epistemological asymmetry between empirical scientific discovery and mathematical arguments. Typically, empirical sciences allow graded evaluation signals–such as prediction error, likelihood or experimental agreement–which enable heuristic search procedures to navigate smooth optimization landscapes. Mathematical reasoning, however, requires exact and strict logical validity: symbolic derivations are either correct or incorrect. Consequently, proof discovery corresponds to a combinatorial search over symbolic sequences in which valid proofs occupy a sparse subset of the hypothesis space. Drawing on ideas from proof complexity theory, we describe a verification–search asymmetry in formal reasoning that helps explain this difference. Identifying this structural yet conceptual distinction indicates that progress in AI-assisted mathematics will likely depend on hybrid architectures that combine generative exploration with rigorous symbolic verification. The introduced verification–search asymmetry offers a conceptual explanation for why AI architectures excel in empirical scientific discovery while remaining comparatively limited in mathematical reasoning.

1 Introduction

Artificial Intelligence (AI)[1] architectures based on large language models and transformer neural network[2] (LLMs) and autonomous agent[3,4] are rapidly transforming scientific research[5–7]. During the past decade or so, machine learning methods have been deployed in a wide spectrum of disciplines within science[8–10], technology[11,12], engineering[13,14] and mathematics (STEM), helping with tasks ranging from prediction of molecular structures[15,16] and material discovery[17] to code synthesis[18] and experimental design. These systems increasingly operate not only as predictive tools but as exploratory engines capable of proposing hypotheses, generating candidate models, and navigating vast combinatorial search spaces.

The success of AI-assisted discovery has been particularly striking in empirical sciences and engineering, where many problems involve complex systems that are difficult to analyze directly. For instance, recently, the authors in[19] used Google Gemini Deep Think Large Language Models that assisted in accelerating quantum-physics-based mathematical exploration, thereby solving an open problem in theoretical physics. Similar advances have emerged in cell-free protein synthesis[16,20,21], scientific engineering design[21–24], where generative models and reinforcement-learning agents have demonstrated the ability to optimize complex systems whose governing dynamics are only partially understood.

However, this remarkable progress reveals an intriguing and subtle asymmetry. Although AI architectures increasingly contribute to discovery in numerous scientific domains (some of them already cited), comparable breakthroughs in mathematics remain rare. Recent advances show that modern AI architectures can solve many structured competition problems, such as those appearing in the International Mathematical Olympiad[25,26]; impressive, these successes occur in respective domains where solution spaces are comparatively constrained. In contrast, research-level mathematical reasoning involves navigating combinatorial expanding spaces of symbolic derivations in which valid proofs occupy an extremely small subset.

Despite significant advances in automated theorem proving[27] and formal verification, current AI architectures seldom produce the kind of deep conceptual insight that characterizes the major developments in contemporary mathematics. This disparity raises a natural and deep question: **why do artificial intelligence architectures appear well suited for discovery in empirical sciences but struggle with the forms of reasoning central to mathematical research and advancement?** This question deserves careful examination. Motivated by this very perspective, we investigate the structural origins of this asymmetry. We argue that the contrast arises not primarily from limitations of the present-day machine learning or robust neural network architecture, but rather from deeper differences between the epistemic structures of scientific discovery and