Evidence-Based Frameworks for Generative Artificial Intelligence

Thomas F. Heston[1,2,*]

1. Department of Family Medicine, University of Washington, Seattle, Washington, USA.

2. Department of Medical Education and Clinical Sciences, Washington State University, Spokane, Washington, USA.

* Corresponding author: Email: theston@uw.edu

Abstract

Generative artificial intelligence and large language models increasingly influence education, research, and decision support. However, current systems often prioritize narrative popularity over evidentiary strength, leading to information cascades and citation loops. This paper proposes a multidisciplinary framework for evaluating AI-generated claims based on the methodological discipline of evidence-based medicine. By integrating blockchain technology for evidence provenance and multi-agent audit protocols, generative systems can transition from narrative-based outputs to evidence-based syntheses. This approach ensures that information is weighted by methodological reliability rather than frequency of repetition, enhancing the trustworthiness of artificial intelligence in high-stakes environments.

Keywords: generative AI, evidence-based medicine, blockchain, information provenance, multi-agent audit, epistemology, decision science

Introduction

Artificial intelligence (AI) has emerged as one of the most influential technologies shaping modern society. Generative AI and large language models (LLMs) now participate in education, research, public discourse, and decision support across multiple fields. Yet the way these systems handle information still relies heavily on a simple statistical fact: claims that appear more often in the training data exert more influence on the model [1]. While useful for language generation, this mechanism is a weak foundation for truth-seeking. Popularity is not synonymous with evidentiary strength, and repetition is not validation [2].

Medicine confronted a similar problem decades ago. Clinical practice was once driven heavily by authority, tradition, and expert opinion. Over time, physicians discovered that widespread belief was often incorrect. Treatments once considered standard later proved ineffective or harmful when subjected to rigorous testing. The response was not cynicism but methodological reform. The medical field was arguably the first to consolidate these ideas into a formal, named discipline, Evidence-Based Medicine (EBM), which provided a hierarchy of evidence [3].

The roots of this transition extend deep into the history of experimental science and mathematical logic. The modern evidence-based framework was built upon the statistical innovations of R.A. Fisher [4], whose work in experimental design and randomization in the early twentieth century provided the mathematical foundation for the randomized controlled trial. This effort was later catalyzed by Archie Cochrane, who famously advocated for the systematic review of all medical interventions, leading to the global research collaboration that bears his name [5]. Furthermore, the integration of Boolean logic and Bayesian analysis—pioneered in a clinical context by researchers such as Alvan Feinstein —allowed clinicians to move beyond simple frequentist statistics to incorporate prior probability and rigorous diagnostic reasoning.

This multidisciplinary methodological reform is uniquely relevant to the development of generative AI. Generative systems require a similarly disciplined framework to distinguish stronger from weaker evidence, weighting sources by methodological reliability rather than narrative dominance. Integrating blockchain technology provides the necessary infrastructure for this evolution by establishing an immutable record of evidence provenance and enabling decentralized multi-agent audits.

The Central Weakness of Narrative-Based Systems

Most LLMs are trained on enormous corpora of text drawn from the internet, books, articles, and public discussions. The statistical training process rewards patterns that occur frequently across the dataset. The more often a claim appears, the more likely the model is to reproduce or reinforce it [6]. This mechanism introduces several well-known problems from epistemology and information science.

First, the majority fallacy suggests that an idea repeated thousands of times is automatically correct. Information cascades occur when writers repeat a claim simply because others have already repeated it. Over time, repetition creates the illusion of consensus. Second, citation loops occur when journalists, commentators, and academics cite one another in ways that make a narrative appear to have many independent sources, even when it originates from a small number of initial claims [7].

The Evidence Hierarchy as a Model for AI

Evidence-based medicine introduced a structured hierarchy that distinguishes the strength of different types of evidence. At the top of this hierarchy are randomized controlled trials and systematic reviews, while observational studies and expert opinion occupy lower levels [8]. The crucial insight is that evidence must be evaluated based on methodological reliability, not simply on frequency or authority.

A single well-designed randomized trial may outweigh hundreds of expert commentaries. This principle directly addresses the problem that AI systems currently face. When language models treat repeated claims as inherently more credible, they risk reinforcing information cascades. Future systems should learn to distinguish primary evidence from commentary and independent confirmation from rhetorical amplification.

Rather than presenting certainty where none exists, evidence-based systems should emphasize probability and risk-benefit analysis, aligning AI development with the mature traditions of reasoning under uncertainty.

Blockchain and Evidence Provenance

A fundamental challenge in the current internet ecosystem is the distortion of claims as they are repeated and summarized, making it difficult to identify the original sources. Blockchain technology offers a solution through evidence provenance. An immutable public ledger can be used to record primary evidence objects, including datasets, trial registrations, and authorship history [9]

The point is not that blockchain determines truth, but that it can preserve provenance, chronology, and source independence. If 10 articles all derive from a single original source, an evidence ledger could make that dependency visible [10]. The Oracle Problem—ensuring data accuracy before it is committed to the ledger—is addressed through decentralized verification. In this model, a network of independent human contributors and autonomous agents performs the validation. Evidence objects are publicly submitted, then reviewed, challenged, and elevated in credibility through transparent, auditable consensus protocols.

Just as blockchain innovation has revolutionized sustainable supply chain management by providing traceability [11,12]. It can do the same for the "supply chain" of information. A blockchain-based evidence ledger creates a verifiable audit trail for every claim an AI system reproduces, allowing users to trace information back to its empirical roots.

Multi-Agent Audit and Epistemic Pluralism

The development of a shared, auditable evidence layer enables a multi-agent environment where competing AI systems can inspect the same underlying materials. We should not want a single generative AI system operating as a solitary epistemic authority. Instead, competing systems should be able to audit shared public records and reach their own syntheses.

This approach promotes epistemic pluralism—the idea that multiple valid interpretations can exist for a shared set of data. Disagreement among AI models would be desirable, mirroring the scientific process in which progress comes from competing interpretations of shared evidence. Such a framework prevents "model monoculture," in which all AI systems converge on a single biased narrative due to nearly identical training datasets.

Blockchain supplies the audit trail, while an evidence-based framework supplies the epistemology. The essential task for generative AI then becomes not parroting the loudest narrative, but optimally synthesizing partially conflicting evidence from a transparent, auditable record.

Case Study: Decentralized Verification in Medical Research

Imagine a decentralized platform where clinical trial data is submitted to a blockchain ledger. Each data point is timestamped and cryptographically signed by the original investigators. When a generative AI system is asked about the efficacy of a specific treatment, it does not merely search the internet for articles; it queries the ledger for primary datasets.

In this case, the AI would be programmed to prioritize the ledger-verified primary data over the statistical echoes of social media or editorial commentary [13]. If a meta-analysis on the ledger indicates high-quality evidence, the AI can more confidently report the finding as stable. If the evidence quality is low, the AI can explicitly recognize the uncertainty.

Practical Implementation and Governance

The hardest governance question is who approves inclusion into the evidence layer. Centralized media or academic gatekeepers cannot be the sole arbiters, as they are part of the current narrative-shaping problem. A more credible model involves decentralized verification, in which evidence is elevated through transparent, peer-reviewed procedures recorded on the blockchain [8].

Conflict of interest must also be visible. Just as medical journals require disclosure of financial relationships, AI systems should incorporate evidentiary weighting that accounts for the incentives of the source generators. The goal is to build an evidence-based system rather than a narrative-based one, ensuring that generative AI becomes a trustworthy partner in human reasoning.

Limitations

Sustainable Infrastructure Funding: The long-term financial cost of maintaining a high-throughput, decentralized evidence ledger remains a significant barrier. Potential solutions include institutional subscription models, tiered submission fees for commercial research, or decentralized grants from public scientific agencies.

Strategic Manipulation: Malicious actors could attempt to skew the evidence base by flooding the ledger with high-volume, low-quality, or fabricated primary data. This risk can be mitigated through cryptographic staking requirements for submitters, reputation-based weighting for verifiers, and multi-agent audit protocols designed to detect and flag anomalous data patterns. A mechanism for the clear identification of corrupted or fraudulent data would result in the permanent identification of bad actors.

Implementation and Integration Complexity: The technical and cultural shift from traditional narrative publication models to a decentralized evidence ledger is immense. A phased implementation strategy—prioritizing high-stakes clinical trial data and developing open-source bridge protocols for existing academic publishers—could facilitate a manageable transition.

Data Privacy and Regulatory Conflict: The immutable nature of blockchain documentation fundamentally conflicts with global "right to be forgotten" mandates, such as the General Data Protection Regulation, which requires the ability to redact or delete retracted or sensitive information [14]. This can be mitigated by storing only cryptographic hashes on the ledger while maintaining primary evidence in erasable off-chain storage, or by using zero-knowledge proofs to verify the evidence without exposing the underlying sensitive data.

Conclusion

The transition from a narrative-based to an evidence-based framework is essential for generative artificial intelligence to become a reliable partner in human reasoning. By adopting the methodological discipline pioneered in medicine and implementing a structured evidence hierarchy, AI systems can move beyond simple statistical pattern completion to prioritize methodological reliability over narrative popularity. Integrating blockchain technology as a decentralized infrastructure for evidence provenance ensures that claims remain auditable, transparent, and grounded in verified primary data rather than distorted citation loops. Ultimately, a system that utilizes multi-agent audits and decentralized verification will foster epistemic pluralism, allowing artificial intelligence to navigate uncertainty with the same rigor and transparency required in high-stakes clinical decision-making.

Acknowledgements

Funding: This study did not receive any external funding.

Conflicts of Interest: The author declares no conflicts of interest.

Data Availability: Not applicable for this conceptual review.

AI Usage: Large language models were used for language editing and formatting assistance; the author reviewed, verified, and is fully responsible for all content.

References

1. Idan D, Einav S: Primer on large language models: an educational overview for intensivists. Crit Care. 2025, 29:238. 10.1186/s13054-025-05479-4

2. Bender EM, Gebru T, McMillan-Major A, Shmitchell S: On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. Association for Computing Machinery: New York, NY, USA; 2021. 610–23.10.1145/3442188.3445922

3. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS: Evidence based medicine: what it is and what it isn’t. BMJ. 1996, 312:71–2. 10.1136/bmj.312.7023.71

4. Fisher RA: The design of experiments. Oliver and Boyd. Lond EdinburghGoogle Sch. 1935.

5. Cochrane AL: Effectiveness and efficiency: random reflections on health services. Nuffield Provincial Hospitals Trust: London; 1972.

6. Kandpal N, Deng H, Roberts A, Wallace E, Raffel C: Large Language Models Struggle to Learn Long-Tail Knowledge. 2023. 10.48550/arXiv.2211.08411

7. Caled D, Silva MJ: Digital media and misinformation: An outlook on multidisciplinary strategies against manipulation. J Comput Soc Sci. 2022, 5:123–59. 10.1007/s42001-021-00118-8

8. Heston TF: The blockchain-based scientific study. Digit Med. 2017, 3:66. 10.4103/digm.digm_17_17

9. Nakamoto S: Bitcoin: A Peer-to-Peer Electronic Cash System.

10. Wang X, Xie H, Ji S, Liu L, Huang D: Blockchain-based fake news traceability and verification mechanism. Heliyon. 2023, 9:e17084. 10.1016/j.heliyon.2023.e17084

11. Manfrino M: Blockchain Innovation for Sustainable Supply Chain Management under EU Corporate Sustainability Reporting Directive (CSRD) Regulation. Int J Blockchain Technol Appl. 2024, 2:. 10.18178/IJBTA.2024.2.1.8-15

12. Seuring S, Müller M: From a literature review to a conceptual framework for sustainable supply chain management. J Clean Prod. 2008, 16:1699–710. 10.1016/j.jclepro.2008.04.020

13. Shumailov I, Shumaylov Z, Zhao Y, Papernot N, Anderson R, Gal Y: AI models collapse when trained on recursively generated data. Nature. 2024, 631:755–9. 10.1038/s41586-024-07566-y

14. European Parliament and Council of the European Union: General Data Protection Regulation (EU) 2016/679. 2016.