Sep 20, 2025

Dear Editor,

We are pleased to submit our manuscript entitled “Estimation of Medical Diagnostic Likelihood Ratios Using Artificial Intelligence” for consideration in NEJM AI.

Bayesian reasoning has long been recognized as the normative standard for diagnostic decision-making, but its application in routine clinical care is limited by a fundamental bottleneck: the scarcity of empirically measured likelihood ratios (LRs). Exhaustive LR measurement is infeasible because diagnostic accuracy studies are costly, slow, and context dependent. As a result, the majority of clinical reasoning still relies on heuristics and gestalt rather than quantitative inference. Our study addresses this barrier by evaluating whether large language models (LLMs) can generate valid, auditable LR estimates that could serve as the missing infrastructure for Bayesian reasoning in clinical care. Using 700 literature-reported LRs from TheNNT.com, we compared outputs from three contemporary LLMs (GPT-4o, o3, GPT-5) and found that modern models approximate empirical values with negligible bias and bounded dispersion. These findings suggest that LLMs can supply the inferred LRs necessary to scale Bayesian reasoning beyond the limited empirical evidence base.

Importantly, this work speaks to whether models can move beyond memorization to genuine prospective inference. We propose that model-generated LRs should function as a priori hypotheses for future diagnostic accuracy studies and as benchmarks for model performance, positioning LLMs as engines for medical discovery. By demonstrating that modern LLMs can approximate diagnostic LRs with auditability and bounded error, our study highlights a path toward AI-assisted Bayesian reasoning that could strengthen clinical diagnosis, hypothesis generation, and the broader practice of evidence-based medicine. We believe these insights will be of strong interest to NEJM AI’s readership.

All authors have declared relevant competing interests

Shuhan He, MD

