

# Dr. CaBot: The AI Expert Discussant

Thomas A Buckley, Riccardo Conci, Peter G Brodeur, Jason Gusdorf, Sourik Beltrán, Bitá Behrouzi, Byron Crowe, Jacob Dockterman, Muzzammil Muhammad, Sarah Ohnigian, Andrew Sanchez, James A Diao, Aashna P Shah, Daniel Restrepo, Eric S Rosenberg, Andrew S Lea, Marinka Zitnik, Scott H Podolsky, Zahir Kanjee, Raja-Elie E Abdounour, Jacob M Koshy, Adam Rodman, Arjun K Manrai

## Background

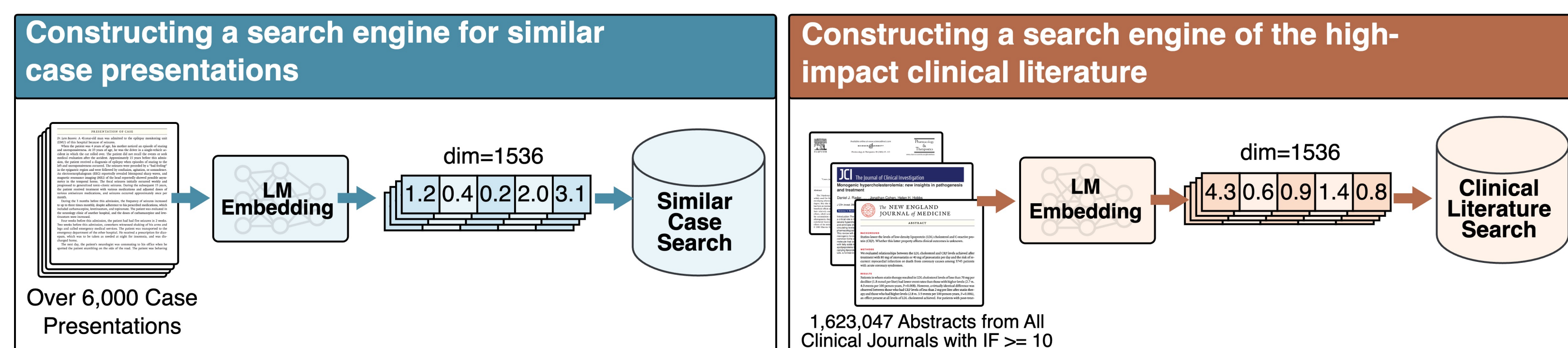
Prior work has shown that LLMs can generate accurate differential diagnoses from complex cases [1, 2].

Unknown if LLMs can:

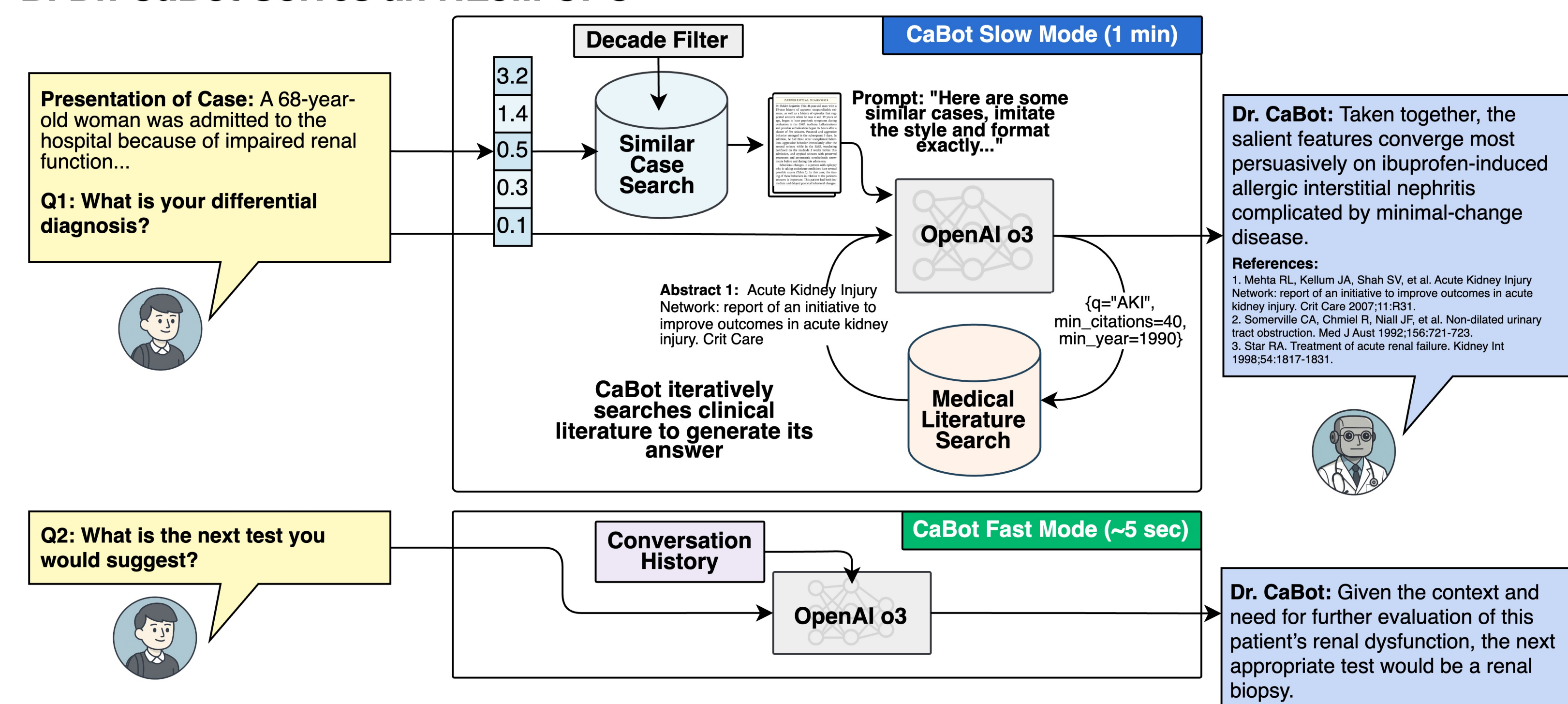
1. Produce convincing, **long-form differential diagnoses** like *NEJM CPCs*.
2. Provide **persuasive clinical reasoning** with supporting literature citations.
3. Present a case in **video format**, as human discussants do.

## Methods

### A. Searching all CPCs and the Clinical Literature



### B. Dr. CaBot Solves an NEJM CPC



### C. Dr. CaBot Presents a Video Case Conference

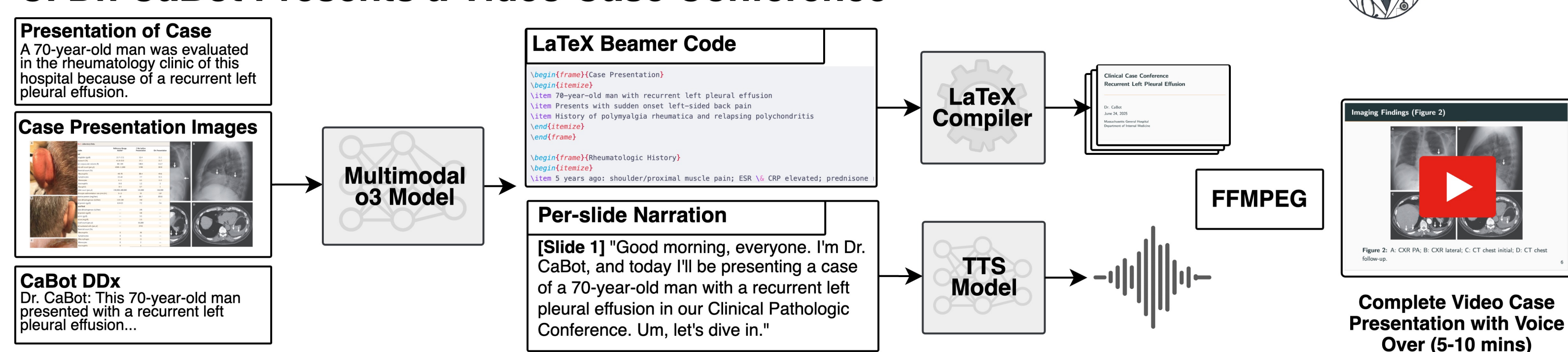


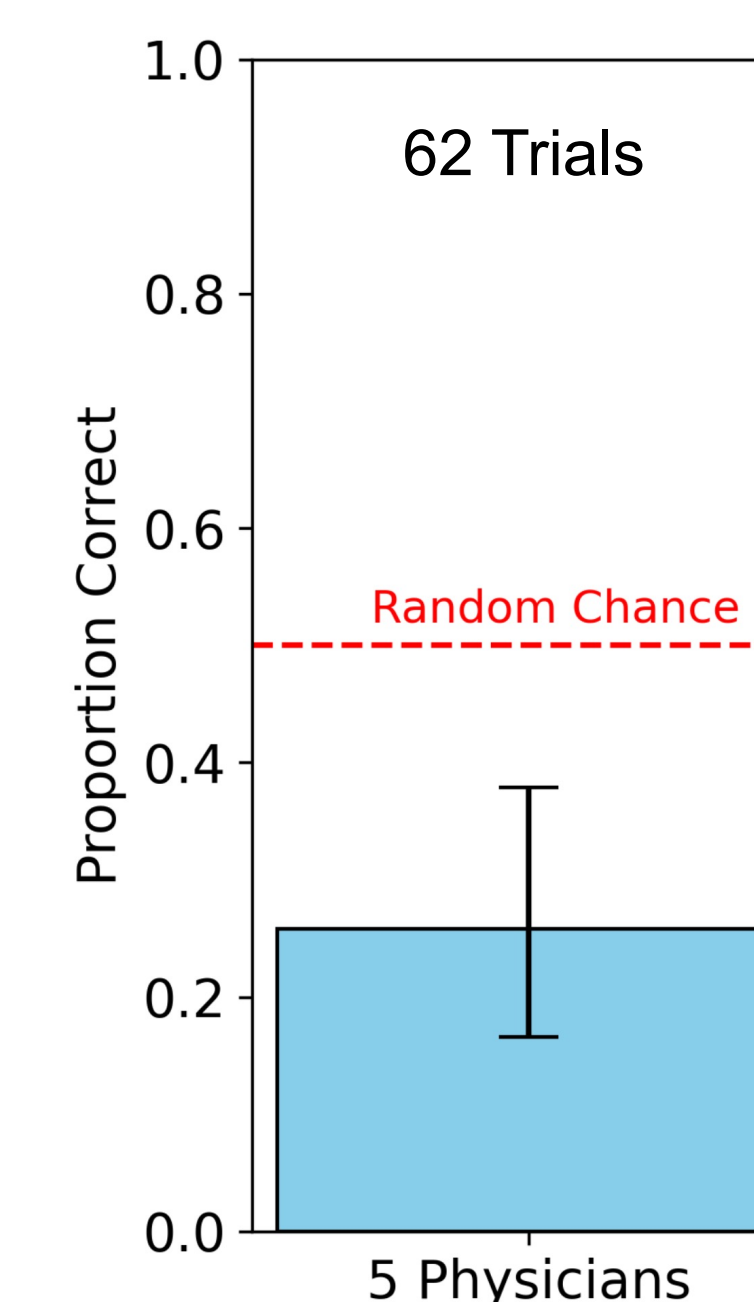
Figure 2: Constructing an AI Discussant

## Results

We assigned 5 internal medicine physicians either:

1. The original expert discussion section published in *NEJM*
2. A discussion section generated by Dr. CaBot

### A. Physician Accuracy in Classifying AI-generated vs Physician-generated Discussion



### B. Physician Quality Ratings

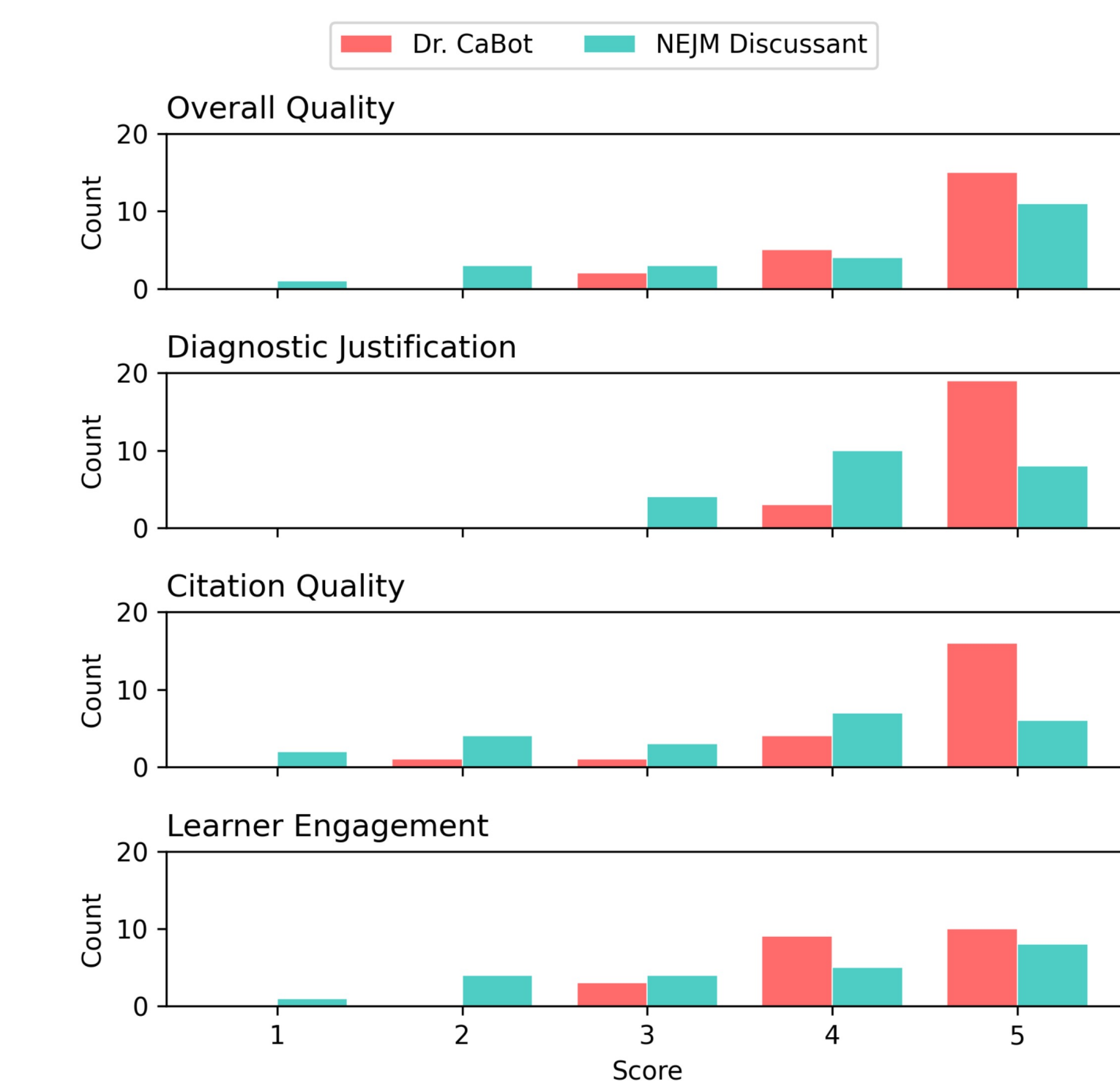


Figure 3: A. Mean accuracy of 5 internal medicine physicians in identifying the source of a differential diagnosis as either AI or physician generated for 62 trials. 95% Confidence intervals generated using Wilson method. B. Distribution of quality ratings for the differential diagnosis published in the CPC compared to those generated by Dr. CaBot. Scores were provided for 44 total differential diagnoses by four internal medicine physicians who were blinded to the source of the differential.

## Go to [cpcbench.com](https://cpcbench.com) to try CaBot, Read the Preprint, and View the Benchmark

**CPC-Bench**  
A Community Benchmark for Clinical AI

Thomas Buckley

Paper Dataset CaBot Submit About

**Overview**

CPC-Bench is a benchmark for evaluating the clinical reasoning capabilities of AI models using expert-curated cases licensed from *The New England Journal of Medicine*. The benchmark consists of 10 distinct tasks that test various aspects of clinical reasoning, from differential diagnosis to medical image interpretation.

- 7,102 Clinical Cases
- 100+ Years of CPCs
- 10 Benchmark Tasks
- 4,009 Unique Diagnoses

**Dr. CaBot: The AI Expert Discussant**

Dr. CaBot is an AI that provides comprehensive differential diagnoses in the style of an expert discussant. Dr. CaBot can search the clinical literature to produce an evidence-based response.

**Clinical Literature Retrieval:** Dr. CaBot searches through 1,623,047 clinical abstracts from leading clinical journals to generate its response.

**Style Adaptation:** The model is provided the two most similar case presentations from the chosen era. These help the model simulate the style of an expert discussant.

**Educational Use Only**  
Dr. CaBot is intended for educational and research purposes only and should not be used for actual medical diagnosis or treatment decisions. This AI tool is designed to simulate clinical reasoning for educational and research scenarios. Always consult qualified healthcare professionals for medical advice, diagnosis, or treatment.

By using this tool, you agree to our [AI Terms of Use](#).

Search cases by title or content...

Case 30-2007: A Newborn Girl with Skin Lesions

Random Custom Text Input

**Case 30-2007: A Newborn Girl with Skin Lesions**

## References

1. Z. Kanjee, B. Brodeur, and A. Rodman, "Accuracy of a Generative Artificial Intelligence Model in a Complex Diagnostic Challenge," *JAMA*, vol. 330, no. 1, pp. 78–80, Jul. 2023.
2. P. G. Brodeur, T. A. Buckley, et al., "Superhuman performance of a large language model on the reasoning tasks of a physician," *arXiv [cs.AI]*, Dec. 14, 2024. [Online]. Available: <https://arxiv.org/abs/2412.10849>
3. T. A. Buckley et al., "Advancing medical artificial intelligence using a century of cases," *arXiv [cs.AI]*, Sep. 15, 2025. [Online]. Available: <https://arxiv.org/abs/2509.12194>