Advancing Artificial Intelligence Using a Century of Cases

Thomas Buckley

Manrai Lab

HARVARD MEDICAL SCHOOL
BLAVATNIK INSTITUTE
BIOMEDICAL INFORMATICS

BRAVE NEW WORLD DEPT.

IF A.I. CAN DIAGNOSE PATIENTS, WHAT ARE DOCTORS FOR?

Large language models are transforming medicine—but the technology comes with side effects.

By Dhruv Khullar

September 22, 2025

BRAVE NEW WORLD DEPT.

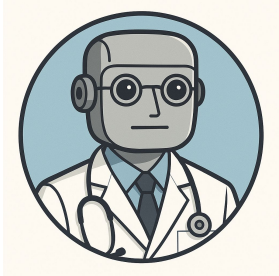
IF A.I. CAN DIAGNOSE PATIENTS, WHAT ARE DOCTORS FOR?

Large language models are transforming medicine—but the technology comes with side effects.

By Dhruv Khullar

September 22, 2025

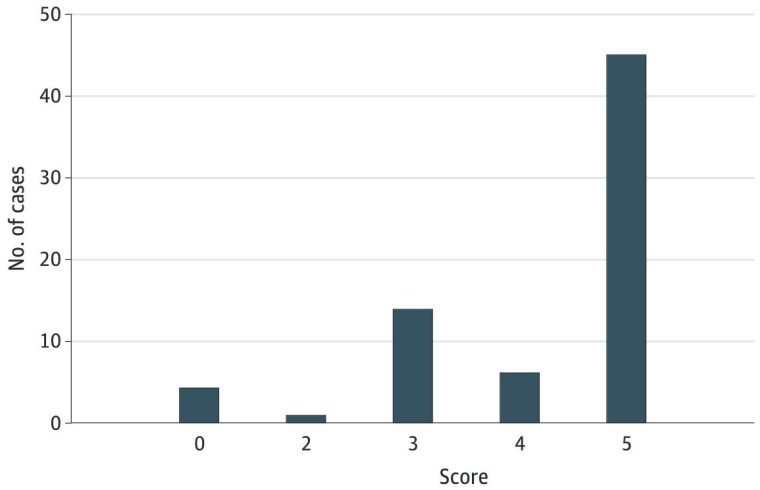
Dr. CaBot



LLMs Solve Complex Clinical Challenges

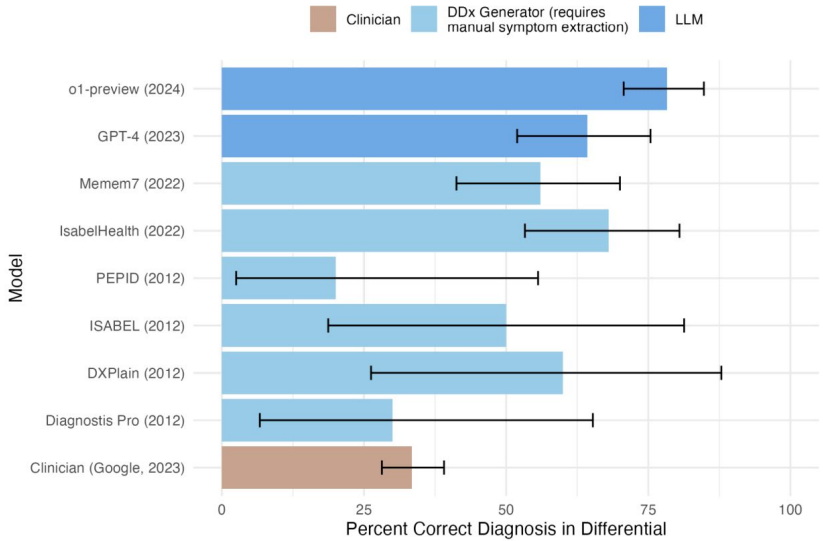
JAMA

2023



70 *NEJM* CPCs, January 2021 to December 2022

2024



143 diagnostic cases from January 2021 to September 2024

1. Kanjee, Z., Crowe, B. & Rodman, A. Accuracy of a Generative Artificial Intelligence Model in a Complex Diagnostic Challenge. JAMA 330, 78–80 (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: http://arxiv.org/abs/2412.10849

Vol. 303 No. 4 CASE RECORDS OF THE MASS.

CASE RECORDS
OF THE
MASSACHUSETTS GENERAL HOSPITAL

Weekly Clinicopathological Exercises

FOUNDED BY RICHARD C. CABOT

ROBERT E. SCULLY, M.D., *Editor*
JAMES J. GALDABINI, M.D., *Associate Editor*
BETTY U. MCNEELY, *Assistant Editor*

CASE 29-1980

https://khuranalab.bwh.harvard.edu/vik-and-claudio-present-new-england-journal-of-medicine-cpc/
https://x.com/nejmcpc

NEJM Clinicopathologic Conference

Case 21-2025: A 75-Year-Old Man with Cough, Dyspnea, and Hypoxemia

Asha M. Anandaiah, M.D.,[1,2] Melissa C. Price, M.D.,[3,4] Jehan W. Alladina, M.D.,[2,5] Pierre O. Ankomah, M.D., Ph.D.,[2,5] and Erik H. Klontz, M.D., Ph.D.[6,7]

PRESENTATION OF CASE

Dr. Jehan W. Alladina: A 75-year-old man was admitted to this hospital because of progressive dyspnea, cough, and hypoxemia.

The patient, who had a history of overlapping asthma and chronic obstructive pulmonary disease, had been in his usual state of health until 7 months before the current presentation, when wheezing, productive cough, and chest tightness developed. He was evaluated by his pulmonologist. Intravenous glucocorticoid therapy was administered in the office, and a tapering course of oral prednisone was prescribed.

Three weeks later, the patient was reevaluated by his pulmonologist because of persistent symptoms. The course of prednisone was extended for 2 weeks, and additional testing was performed. The absolute eosinophil count was 910 cells per microliter (reference value, <400), and the IgE level was 4440 IU per milliliter (reference value, <114); a diagnosis of allergic bronchopulmonary aspergillosis was considered. Treatment with dupilumab twice monthly was started.

Six weeks later, and 5 months before the current presentation, fever and chest congestion developed. Chest radiography reportedly revealed opacities in both lower lobes; a diagnosis of community-acquired pneumonia was considered. Treatment with azithromycin and cefpodoxime was started, along with a brief course of intravenous glucocorticoids, followed by treatment with prednisone. *Haemophilus influenzae* grew in a sputum culture. Azithromycin and cefpodoxime were discontinued, and levofloxacin was started. Prednisone was stopped after 2 weeks.

Two months later, and 3 months before the current presentation, the patient was again evaluated by his pulmonologist. He reported ongoing wheezing and chest congestion. The absolute eosinophil count was 100 cells per microliter, the IgE level 516 IU per milliliter, and the *Aspergillus fumigatus* IgE level 27.50 IU per milliliter (reference value, <4.00). Intravenous glucocorticoids were administered, and a tapering course of oral prednisone was prescribed. Treatment with voriconazole was started, with a plan for an 8-week course.

During the 5 weeks before the current presentation, the patient visited an urban area in Central America. While he was there, dyspnea with productive cough devel-

Table 1. Laboratory Data.*

Variable	Reference Range, Adults, This Hospital†	On Presentation, Emergency Department	Evening of First Hospital Day, ICU	Morning of Second Hospital Day, ICU
Hemoglobin (g/dl)	13.5–17.5	13.5	12.7	11.4
Hematocrit (%)	41.0–53.0	39.9	37.5	34.5
White-cell count (per μl)	4500–11,000	7050	6200	4910
Differential count (per μl)				
Neutrophils	1800–7700	6020	5790	4140
Lymphocytes	1000–4800	460	200	470
Monocytes	200–1200	170	110	130
Eosinophils	0–900	—	50	40
Metamyelocytes (%)	0	5.7	0.8	2.6
Platelet count (per μl)	150,000–400,000	227,000	202,000	168,000
Sodium (mmol/liter)	135–145	131	132	134
Potassium (mmol/liter)	3.4–5.0	4.4	5.1	5.2
Chloride (mmol/liter)	98–108	94	99	100
Carbon dioxide (mmol/liter)	23–32	25	24	26
Urea nitrogen (mg/dl)	8–25	19	16	12
Creatinine (mg/dl)	0.60–1.50	1.02	0.84	0.56
Glucose (mg/dl)	70–110	100	126	104
Albumin (g/dl)	3.3–5.0	2.6	2.3	2.0
Lactate (mmol/liter)	0.5–2.0	2.4	1.7	1.6
Phosphorus (mg/dl)	2.6–4.5	—	1.7	2.2
Venous blood pH	7.30–7.40	—	7.44	7.46

* To convert the values for urea nitrogen to millimoles per liter, multiply by 0.357. To convert the values for creatinine to micromoles per liter, multiply by 88.4. To convert the values for glucose to millimoles per liter, multiply by 0.05551. To convert the values for lactate to milligrams per deciliter, divide by 0.1110. To convert the values for phosphorus to millimoles per liter, multiply by 0.3229. ICU denotes intensive care unit.
† Reference values are affected by many variables, including the patient population and the laboratory methods used. The ranges used at Massachusetts General Hospital are for adults who are not pregnant and do not have medical conditions that could affect the results. They may therefore not be appropriate for all patients.

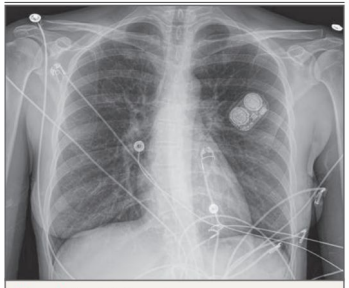


Figure 1. Chest Radiograph.
A single-view frontal image obtained with a portable device shows no evidence of pneumonia or pulmonary edema and no pleural effusion. The heart size is normal. An implantable loop recorder projects over the left hemithorax. Electrocardiographic leads can also be seen.

NEJM Clinicopathologic Conference

Differential Diagnosis

Dr. Asha M. Anandaiah: This 75-year-old man with severe asthma complicated by allergic bronchopulmonary aspergillosis and prolonged exposure to glucocorticoids presented with acute respiratory failure. CT findings included a cavitary lesion in the right upper lobe and mass lesions in both lower lobes. Cavitary and mass lesions in the lung can be caused by inflammatory disorders, cancer, and infections.

VASCULITIS

It would be unusual for a progressive inflammatory disorder to develop during treatment with glucocorticoids, but it is important to consider inflammatory disorders that would be treated with more potent immunosuppressive therapy. Pulmonary vasculitis, specifically granulomatosis with polyangiitis, can result in cavitary and mass lesions, and glucocorticoid monotherapy would be inadequate for treating this condition.[1] However, granulomatosis with polyangiitis is often associated with involvement of the kidneys or other organs, which was not seen in this patient. In addition, well-circumscribed large mass lesions would be an uncommon feature of vasculitis.

CANCER

Patients with lung cancer can present with cavitary and mass lesions. Pneumonic-type lung adenocarcinoma[2] and primary pulmonary lymphoma[3] are important considerations in a patient with large mass lesions. However, the subacute onset of this patient's symptoms and the absence of clinically significant lymphadenopathy on examination and imaging studies make cancer unlikely overall.

INFECTION

The patient's subacute presentation and prolonged exposure to glucocorticoids make infection a likely possibility. His travel history expands the range of possible infections that need to be considered. In working toward the most likely diagnosis, I will focus on pulmonary infections that can result in nodules or masses and cavitary lesions. Although severe or necrotizing bacterial infections can cause both cavitary lesions and large consolidations, this patient's large mass lesions, along with his subacute presentation and prolonged exposure to glucocorticoids, make an opportunistic infection more likely than a typical bacterial infection.

NEJM Clinicopathologic Conference

Fungal Infection

Fungal infections often result in nodules and masses in the lungs. This patient's extensive travel history, including recent travel in North and Central America, puts him at risk for an acute or reactivated endemic fungal infection such as histoplasmosis, blastomycosis, coccidioidomycosis, or paracoccidioidomycosis.[4,5] All these infections are caused by dimorphic fungi that grow as mold in soil in

Parasitic Infection

Parasitic infection should be considered in this patient with an extensive travel history and prolonged exposure to glucocorticoids. Treatment with dupilumab, which blocks the interleukin-4α receptor, may also increase susceptibility to parasitic infection. However, this patient did not have the gastrointestinal symptoms and multisystem involvement that are typical of parasitic disease. Paragonimiasis can

Mycobacterial Infection

The patient had traveled to areas in which *M. tuberculosis* is endemic and subsequently presented with a cavitary lesion in the right upper lobe, a finding that is characteristic of reactivated tuberculosis.[12] Although the interferon-γ release assay performed for the detection of latent *M. tuberculosis* infection was negative, this test has limited sensitivity and cannot be used to rule out active tuberculosis.[13] Infection with nontuberculous mycobacteria can also cause cavitary

Gram-Positive Bacterial Infection

Rhodococcus equi, actinomyces, and nocardia are gram-positive bacteria that can cause cavitary pneumonia in immunocompromised patients. *R. equi*, an opportunistic pathogen that is most often associated with HIV infection, is a pleomorphic gram-positive coccobacillus found in the intestinal tract of herbivores. It is classically acquired through contact with foals or their manure or through inhalation of infected dust particles. The diagnosis of rhodococcus pneumonia often requires a high index of suspicion because it can be mistaken for infection with a contaminating diphtheroid from the oropharynx on the basis of morphologic features or mistaken for infection with a mycobacterium on the basis of positive staining for partially acid-fast bacilli.[15]

NEJM Clinicopathologic Conference

Fungal Infection

Fungal infections often result in nodules and masses in the lungs. This patient's extensive travel history, including recent travel in North and Central America, puts him at risk for an acute or reactivated endemic fungal infection such as histoplasmosis, blastomycosis, coccidioidomycosis, or paracoccidioidomycosis.[4,5] All these infections are caused by dimorphic fungi that grow as mold in soil in

Parasitic Infection

Parasitic infection should be considered in this patient with an extensive travel history and prolonged exposure to glucocorticoids. Treatment with dupilumab, which blocks the interleukin-4α receptor, may also increase susceptibility to parasitic infection. However, this patient did not have the gastrointestinal symptoms and multisystem involvement that are typical of parasitic disease. Paragonimiasis can

Cases have 20-30 citations to the clinical literature

Mycobacterial Infection

The patient had traveled to areas in which *M. tuberculosis* is endemic and subsequently presented with a cavitary lesion in the right upper lobe, a finding that is characteristic of reactivated tuberculosis.[12] Although the interferon-γ release assay performed for the detection of latent *M. tuberculosis* infection was negative, this test has limited sensitivity and cannot be used to rule out active tuberculosis.[13] Infection with nontuberculous mycobacteria can also cause cavitary

Gram-Positive Bacterial Infection

Rhodococcus equi, actinomyces, and nocardia are gram-positive bacteria that can cause cavitary pneumonia in immunocompromised patients. *R. equi*, an opportunistic pathogen that is most often associated with HIV infection, is a pleomorphic gram-positive coccobacillus found in the intestinal tract of herbivores. It is classically acquired through contact with foals or their manure or through inhalation of infected dust particles. The diagnosis of rhodococcus pneumonia often requires a high index of suspicion because it can be mistaken for infection with a contaminating diphtheroid from the oropharynx on the basis of morphologic features or mistaken for infection with a mycobacterium on the basis of positive staining for partially acid-fast bacilli.[15]

Emulating the Expert Discussant

Search the Clinical Literature

Emulating the Expert Discussant

Search the Clinical Literature

Produce a Comprehensive DDx with Clinical Reasoning

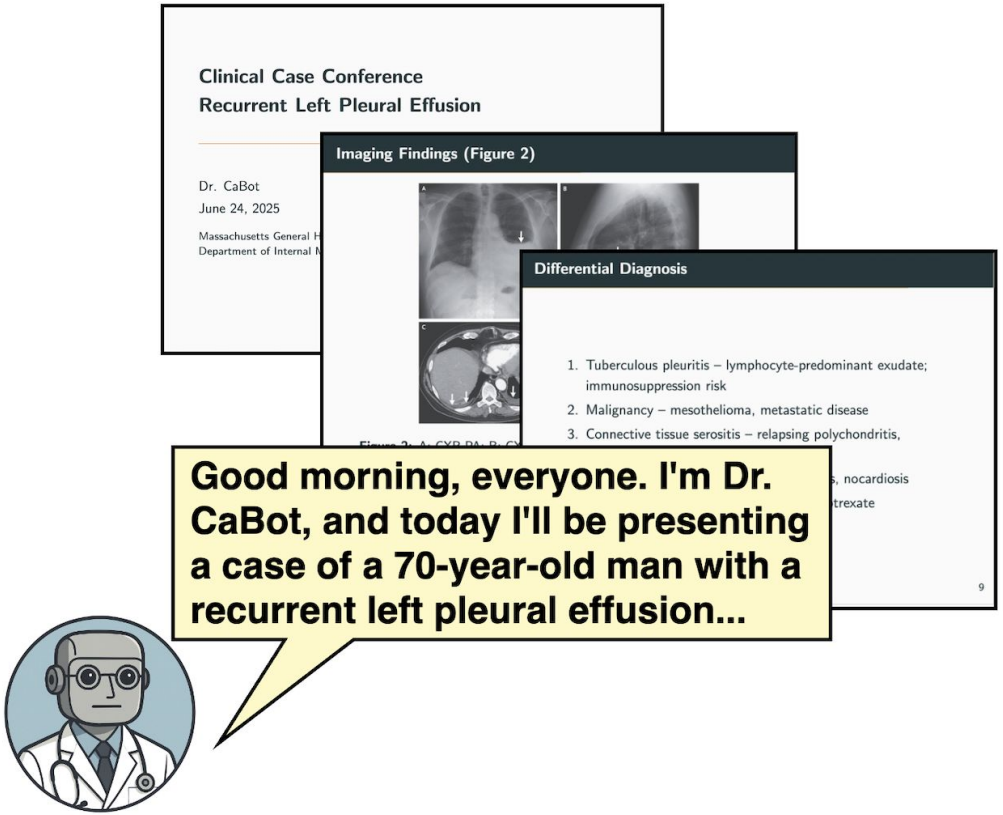
Emulating the Expert Discussant

Search the Clinical Literature

Produce a Comprehensive DDx with Clinical Reasoning

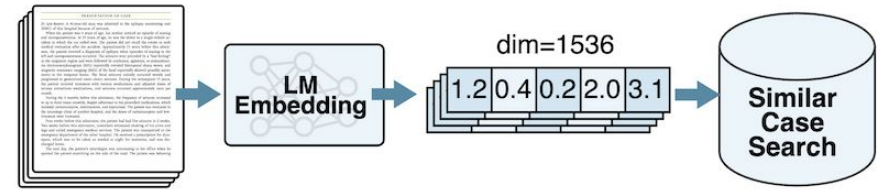
Emulate the Style of the Expert Discussant

Introducing Dr. CaBot: The AI Expert Discussant



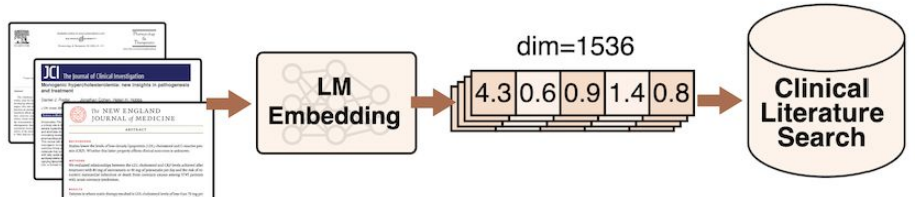
A. Searching all CPCs and the Clinical Literature

Constructing a search engine for similar case presentations



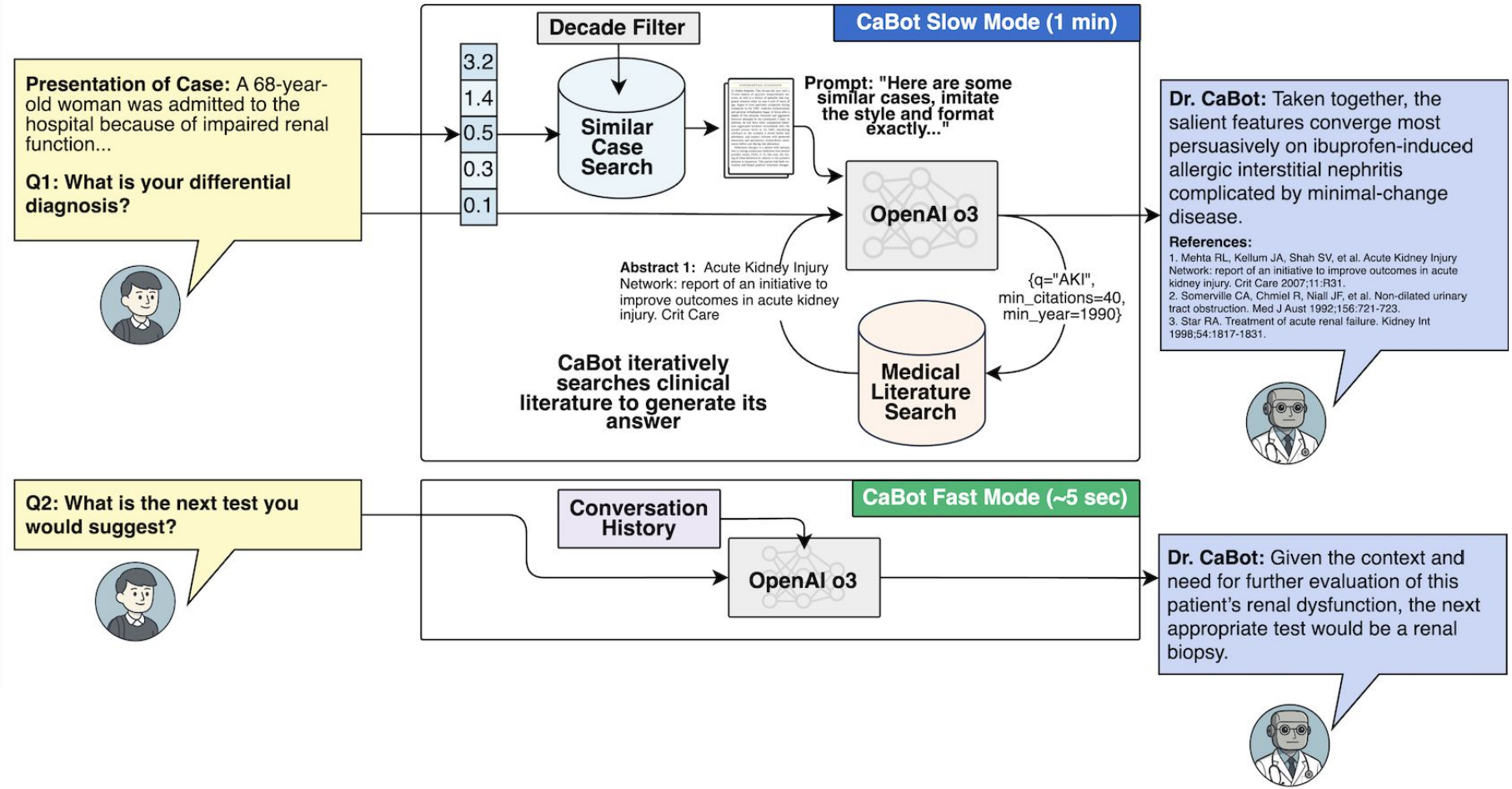
Over 6,000 Case Presentations

Constructing a search engine of the high-impact clinical literature



1,623,047 Abstracts from All Clinical Journals with IF >= 10

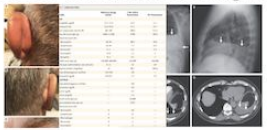
B. Dr. CaBot Solves an NEJM CPC



C. Dr. CaBot Presents a Video Case Conference

Presentation of Case
A 70-year-old man was evaluated in the rheumatology clinic of this hospital because of a recurrent left pleural effusion.

Case Presentation Images



CaBot DDx
Dr. CaBot: This 70-year-old man presented with a recurrent left pleural effusion...

Multimodal o3 Model

LaTeX Beamer Code

```
\begin{frame}{Case Presentation}
\begin{itemize}
\item 70-year-old man with recurrent left pleural effusion
\item Presents with sudden onset left-sided back pain
\item History of polymyalgia rheumatica and relapsing polychondritis
\end{itemize}
\end{frame}

\begin{frame}{Rheumatologic History}
\begin{itemize}
\item 5 years ago: shoulder/proximal muscle pain; ESR \& CRP elevated; prednisone
```

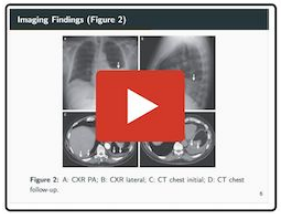
Per-slide Narration

[Slide 1] "Good morning, everyone. I'm Dr. CaBot, and today I'll be presenting a case of a 70-year-old man with a recurrent left pleural effusion in our Clinical Pathologic Conference. Um, let's dive in."

LaTeX Compiler

TTS Model

FFMPEG



Complete Video Case Presentation with Voice Over (5-10 mins)

When Low Sugar Strikes
A 49-Year-Old Man with Recurrent Hypoglycemia

Dr. CaBot

August 10, 2025

Harvard Medical School

When Low Sugar Strikes
A 49-Year-Old Man with Recurrent Hypoglycemia

Dr. CaBot

August 10, 2025

Harvard Medical School

Evaluation

Assigned 5 internal medicine physicians either:

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

Evaluation

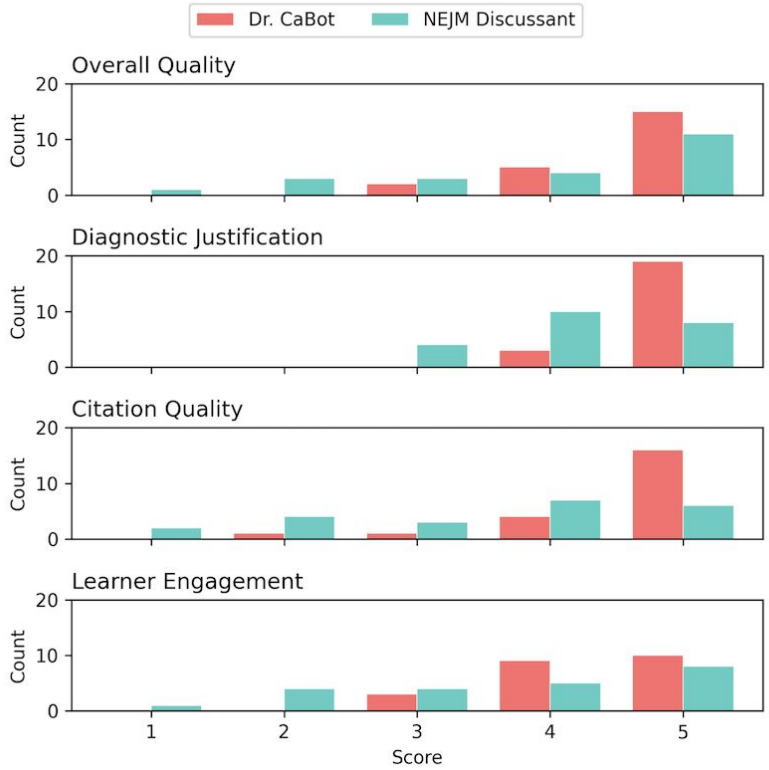
Assigned 5 internal medicine physicians either:

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

In a blinded comparison **physicians misclassified the source of the differential in 46 of 62 (74%) of trials.**

Evaluation

CaBot rated as **overall higher quality** across 44 trials.



cpcbench.com

CPC-Bench

A Community Benchmark for Clinical AI

Thomas Buckley

Paper Dataset CaBot Submit About

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 1.6M+ 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 32-2010: A Pregnant Woman with Abdominal Pain and Fluid i

Random

Custom Text Input

Case 32-2010: A Pregnant Woman with Abdominal Pain and Fluid in the Peritoneal Cavity

Computer Science > Artificial Intelligence

[Submitted on 15 Sep 2025]

Advancing Medical Artificial Intelligence Using a Century of Cases

Thomas A. Buckley, Riccardo Conci, Peter G. Brodeur, Jason Gusdorf, Sourik Beltrán, Bita 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. Abdulnour, Jacob M. Koshy, Adam Rodman, Arjun K. Manrai

BACKGROUND: For over a century, the New England Journal of Medicine Clinicopathological Conferences (CPCs) have tested the reasoning of expert physicians and, recently, artificial intelligence (AI). However, prior AI evaluations have focused on final diagnoses without addressing the multifaceted reasoning and presentation skills required of expert discussants.

METHODS: Using 7102 CPCs (1923-2025) and 1021 Image Challenges (2006-2025), we conducted extensive physician annotation and automated processing to create CPC-Bench, a physician-validated benchmark spanning 10 text-based and multimodal tasks, against which we evaluated leading large language models (LLMs). Then, we developed "Dr. CaBot," an AI discussant designed to produce written and slide-based video presentations using only the case presentation, modeling the role of the human expert in these cases.

RESULTS: When challenged with 377 contemporary CPCs, o3 (OpenAI) ranked the final diagnosis first in 60% of cases and within the top ten in 84% of cases, outperforming a 20-physician baseline; next-test selection accuracy reached

Acknowledgements

Riccardo Conci, Peter G Brodeur, Jason Gusdorf, Sourik Beltrán, Bita 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 Abdulnour, Jacob M Koshy, Adam Rodman, Arjun K Manrai

Preprint

Website