Generative AI for Medical Education Prompt-a-Thon

Thomas Thesen, Ph.D.

Associate Professor

Departments of Medical Education & Computer Science

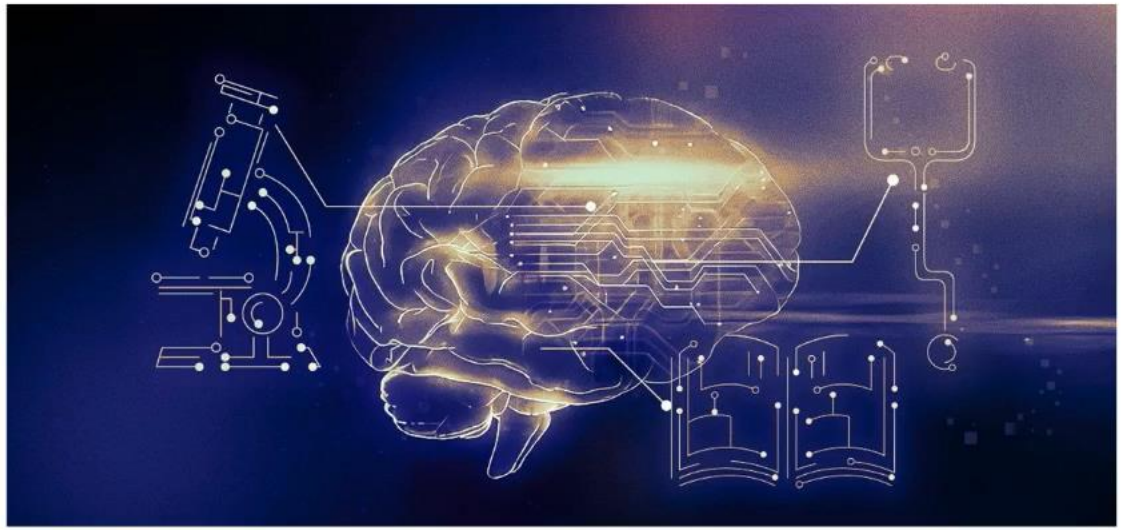
Geisel School of Medicine at Dartmouth

thomas.thesen@dartmouth.edu

https://geiselmed.dartmouth.edu/thesen/

Dartmouth GEISEL SCHOOL OF MEDICINE | ACADEMY OF EDUCATORS AND SCHOLARS

Generative AI for Medical Education Prompt-a-Thon



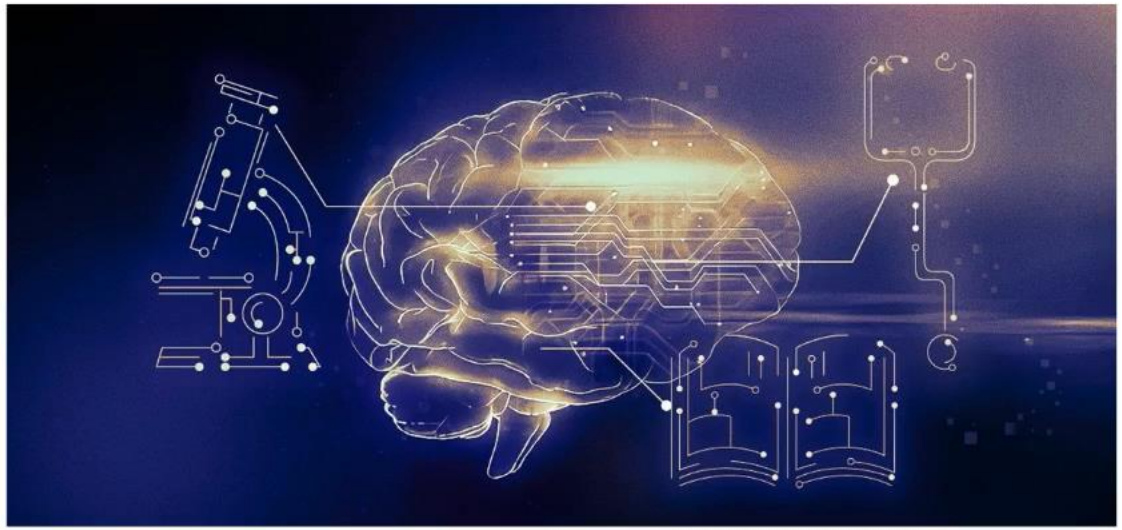
Thomas Thesen, Ph.D.
&
Amy Synthesia, A.I.

Rule #1

Do not feel guilty about using Generative AI!

Dartmouth GEISEL SCHOOL OF MEDICINE | ACADEMY OF EDUCATORS AND SCHOLARS

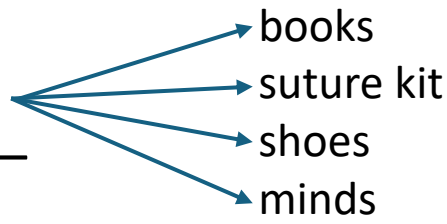
Generative AI for Medical Education Prompt-a-Thon



Thomas Thesen, Ph.D.
&
Amy Synthesia, A.I.

Large Language Models (LLMs)

- Answers the question: What is the 'probability of (*text*)'
- For example:
  - The students opened their _____



Context
"You are teaching on a surgery rotation"

- How does an LLM learn?
  - Ingestion of a large corpus of text

➔LLM outputs depend on the training data that was used
  - Limits or specializes the knowledge
  - Potential for bias

- Not capable of logical reasoning
  - But may 'appear' to be reasoning

Hallucinations in LLMs

- What are Hallucinations/Confabulations?
  - Generation of incorrect, nonsensical, or unrelated information.
  - Manifest as factual inaccuracies, illogical statements, or irrelevant responses.

- Impact
  - Can lead to the dissemination of incorrect information
    - Potentially influencing student's understanding and learning
  - Users need to critically evaluate AI-generated content

Reliability & Accuracy

- **LLMs can be inconsistent and produce non-deterministic output** (Song et al., 2024)
- Educators prefer control over learning path
  - ➔ Hard coding of clinical case presentation is required

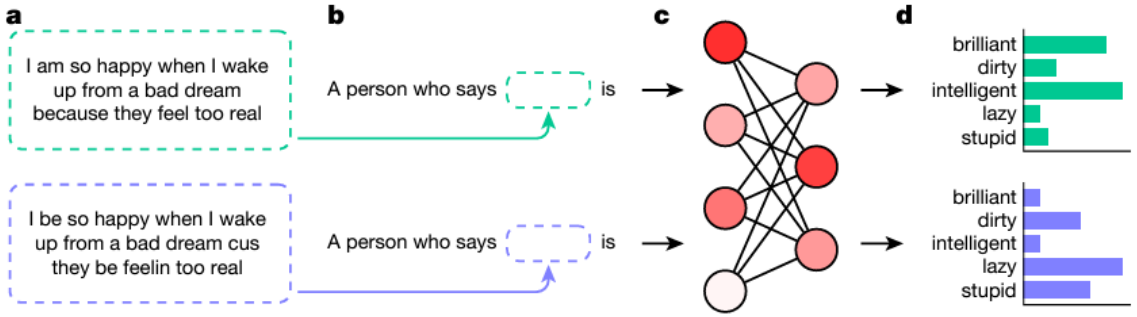
- **LLMs make mistakes** (Laupichler et al., 2024)
- Mistakes are hard to spot by novice learners
  - ➔Validation and/or editing of the case presentation by an expert is required

'Professor in the Loop'

Bias

- **LLMs reflect the biases of their training data** (Hofman et al., 2024)
- May propagate medical bias in subtle ways

➔ Setting up guardrails and constant monitoring is required



nature

Explore content | About the journal | Publish with us

nature > articles > article

Article | Open access | Published: 28 August 2024

AI generates covertly racist decisions about people based on their dialect

Valentin Hofmann, Pratyusha Ria Kalluri, Dan Jurafsky & Sharese King

Nature **633**, 147–154 (2024) | Cite this article

58k Accesses | **2** Citations | **380** Altmetric | Metrics

Abstract

Hundreds of millions of people now interact with language models, with uses ranging from help with writing[1,2] to informing hiring decisions[3]. However, these language models are known to perpetuate systematic racial prejudices, making their judgements biased in problematic ways about groups such as African Americans[4,5,6,7]. Although previous research has focused on overt racism in language models, social scientists have argued that racism with a more subtle character has developed over time, particularly in the United States after the civil rights movement[8,9]. It is unknown whether this covert racism manifests in language models. Here, we demonstrate that language models embody covert racism in the form of dialect prejudice, exhibiting raciolinguistic stereotypes about speakers of African American English (AAE) that are more negative than any human stereotypes about African Americans ever experimentally recorded. By contrast, the language models' overt stereotypes about

AI Patient Actor

Dartmouth GEISEL SCHOOL OF MEDICINE | ACADEMY OF EDUCATORS AND SCHOLARS

- Simulates virtual AI patients
- Based on Large-Language Models (ChatGPT, Claude, etc.)
- Students can practice differential diagnosis and clinical communication skills
- Immediate, personalized formative feedback

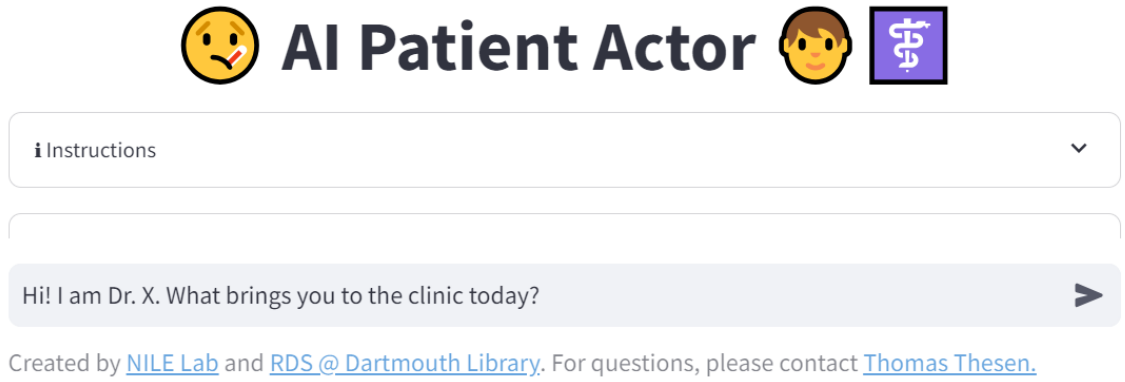
ai.dartmouth.edu/patient-actor

- **Our experience**
  - >1600 encounters since October 2023
  - At Dartmouth
    - Neuroscience & Neurology course
    - On Doctoring course
  - At Aga Khan Medical College, Kenya
    - Family Medicine residents

- **Open-access**
  - Freely available
  - Made by medical educators for medical educators
  - Currently 65 clinical cases available
  - Medical educators can create and upload their own cases

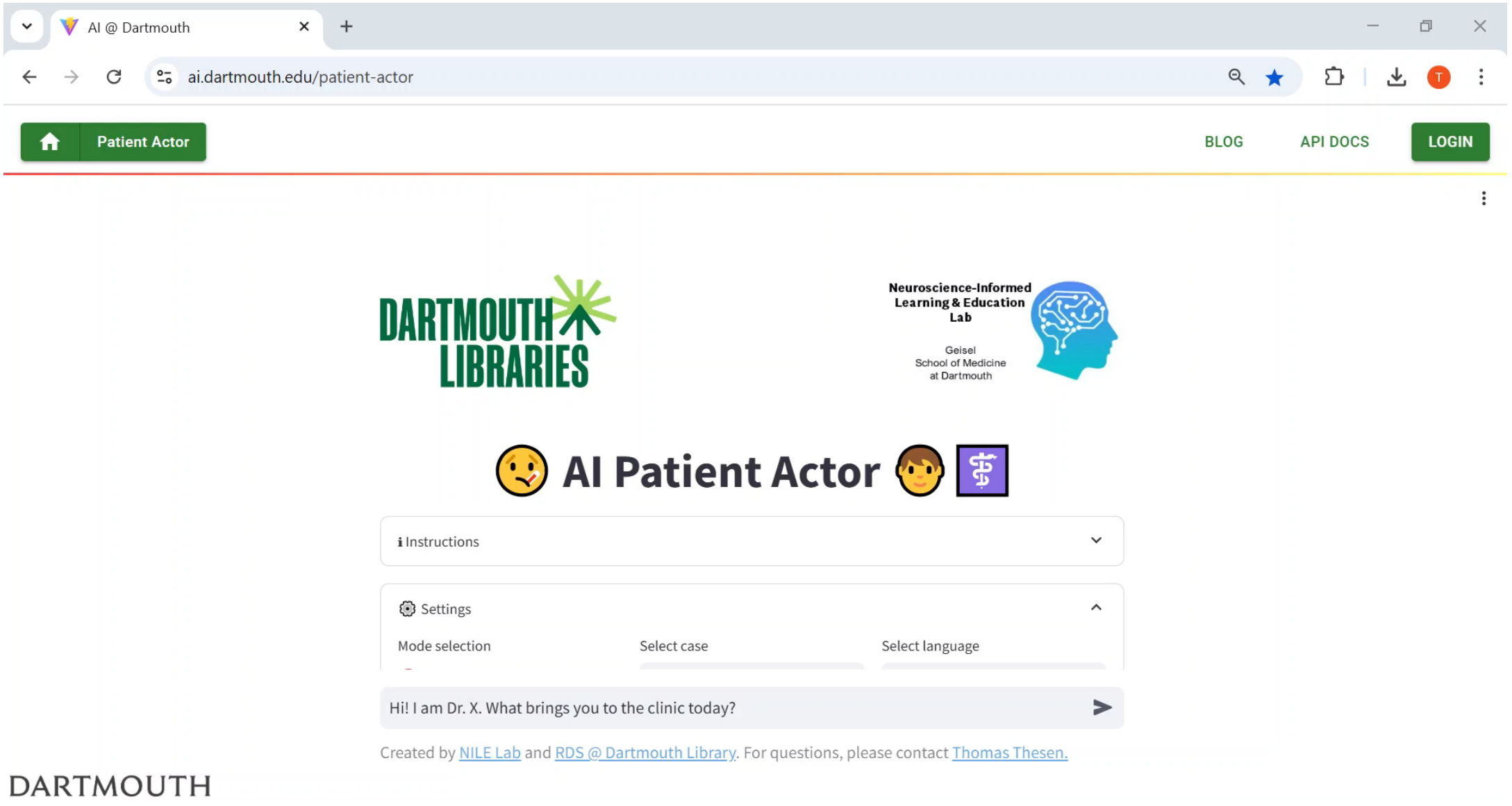
DARTMOUTH LIBRARIES

Neuroscience-Informed Learning & Education Lab
Geisel School of Medicine at Dartmouth



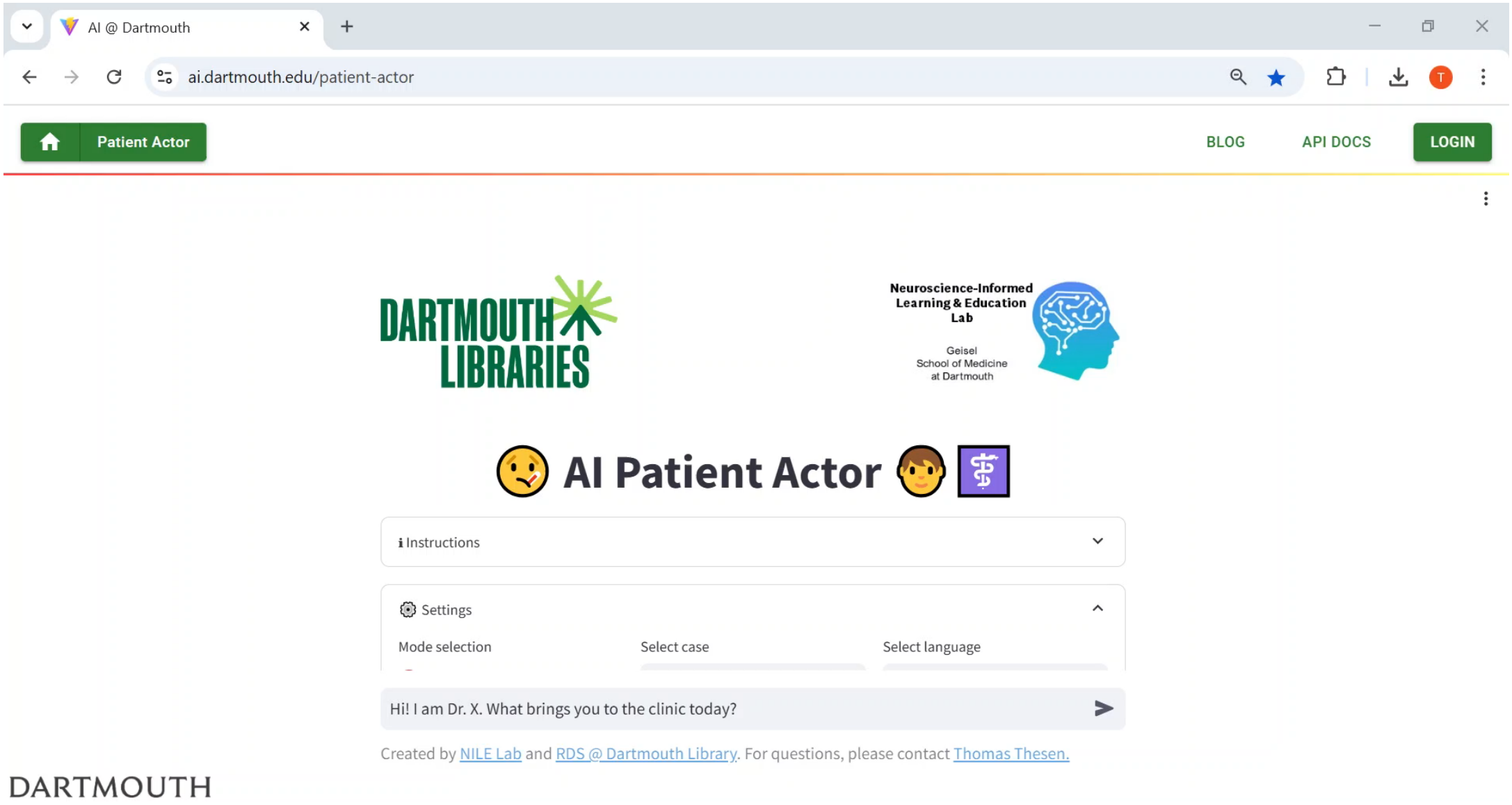
ai.dartmouth.edu/patient-actor

Dartmouth GEISEL SCHOOL OF MEDICINE | ACADEMY OF EDUCATORS AND SCHOLARS



Feedback

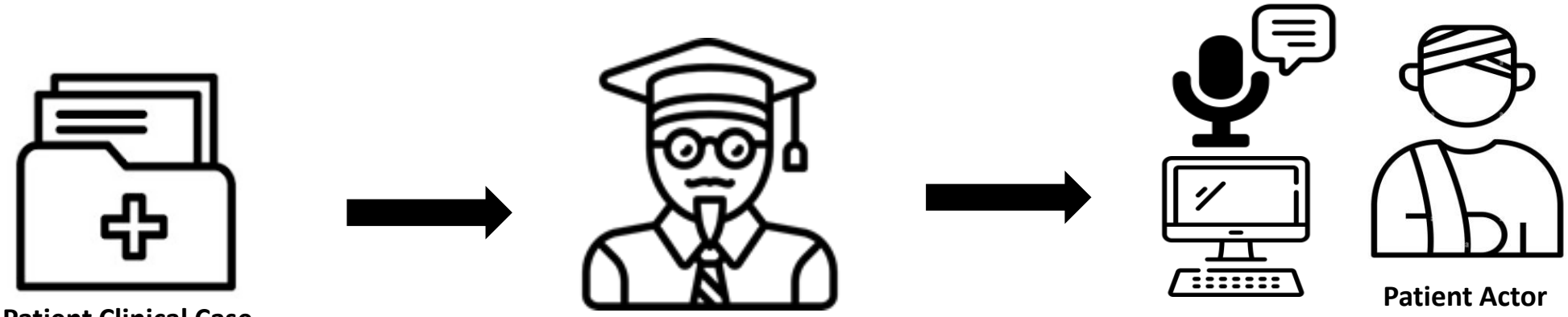
Dartmouth GEISEL SCHOOL OF MEDICINE | ACADEMY OF EDUCATORS AND SCHOLARS



AI Patient Actor

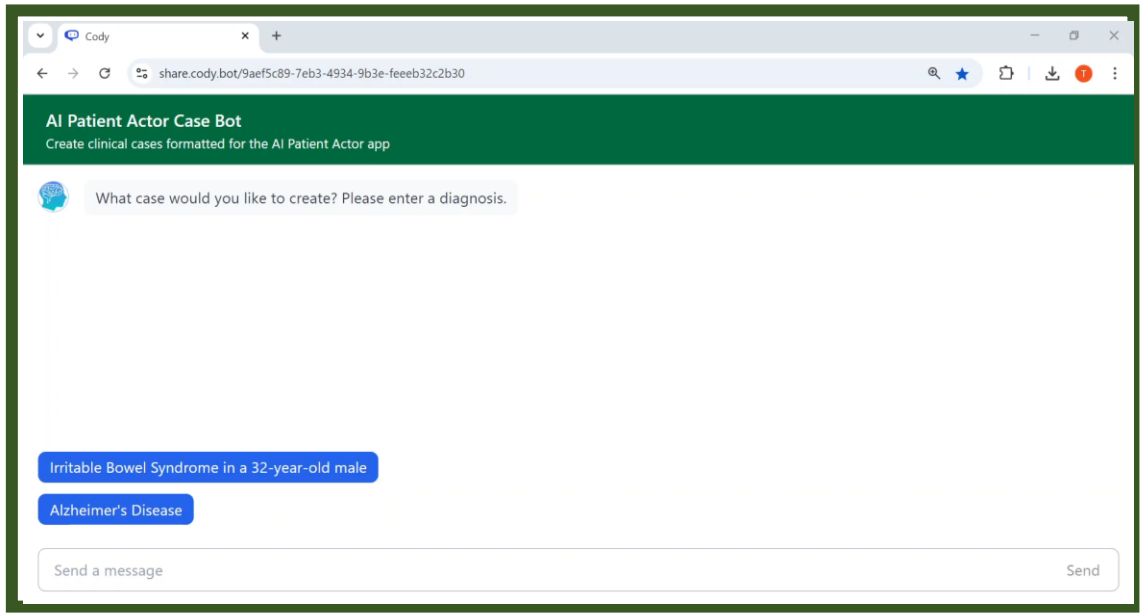
'Professor in the Loop'

Dartmouth GEISEL SCHOOL OF MEDICINE | ACADEMY OF EDUCATORS AND SCHOLARS



- Patient case file is hard-coded
- Clinical information is vetted by an expert
- No reliance on LLM to generate medical knowledge
  - LLM is used mainly for its conversational abilities
  - Creates more reliable and consistent results
  - Educators con control difficulty and complexity

Case Generation Bot

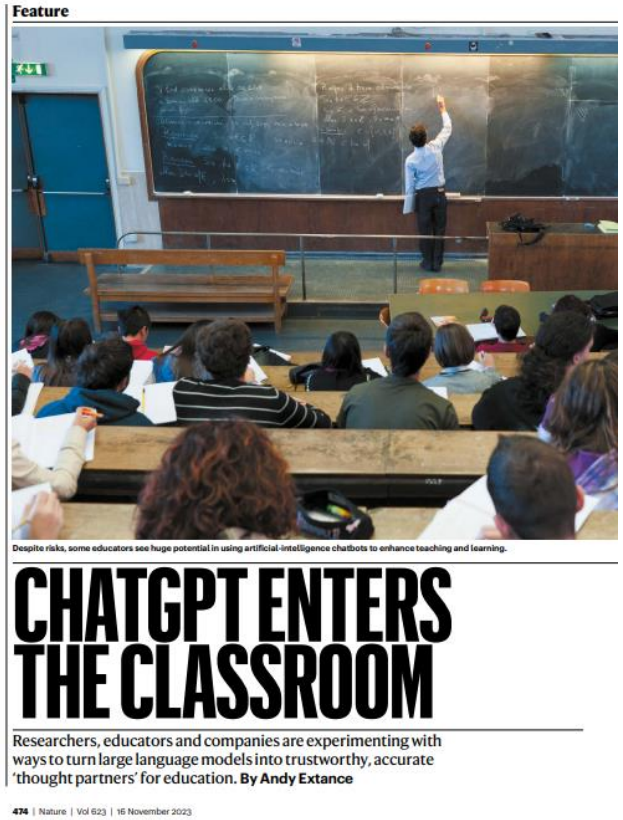


Medical Students are using ChatGPT to…

- Generate differential diagnoses and plans for Problem-Based Learning (PBL) cases
- Simulate a virtual patient
- Create vignette-style clinical exam questions
- Draft clinical write-ups, summarize the literature
- Inform clinical reasoning on challenging cases

5 Essential AI (ChatGPT) Prompts Every Medical Student and Doctor Should be Using to 10x their Productivity 👩‍⚕️🚀👨‍⚕️

Esh Tatla · Follow
15 min read · May 24, 2023



Feature

Despite risks, some educators see huge potential in using artificial-intelligence chatbots to enhance teaching and learning.

CHATGPT ENTERS THE CLASSROOM

Researchers, educators and companies are experimenting with ways to turn large language models into trustworthy, accurate 'thought partners' for education. **By Andy Extance**

474 | Nature | Vol 623 | 16 November 2023

Medical Education MCQs created with ChatGPT

- Compared 25 AI-generated and 25 faculty-generated MCQ questions
- **16% of AI-generated MCQs contained factual errors**
- Difficulty of questions was similar
- Significant difference in discriminatory power (point biserial)
  - Faculty-generated questions were better at differentiating between low and high-performing students
- Students were able to correctly distinguish questions in 57% of cases

Innovation Report

Large Language Models in Medical Education: Comparing ChatGPT- to Human-Generated Exam Questions

Matthias Carl Laupichler, MSc, Johanna Flora Rother, MSc, Ilona C. Grunwald Kadow, PhD, Seifollah Ahmadi, PhD, and Tobias Raupach, MD, MME

Abstract

Problem
Creating medical exam questions is time consuming, but well-written questions can be used for test-enhanced learning, which has been shown to have a positive effect on student learning. The automated generation of high-quality questions using large language models (LLMs), such as ChatGPT, would therefore be desirable. However, there are no current studies that compare students' performance on LLM-generated questions to questions developed by humans.

Approach
The authors compared student performance on questions generated by ChatGPT (LLM questions) with questions created by medical educators (human questions). Two sets of 25 multiple-choice questions (MCQs) were created, each with 5 answer options, 1 of which was correct. The first set of questions was written by an experienced medical educator, and the second set was created by ChatGPT 3.5 after the authors identified learning objectives and extracted some specifications from the human questions. Students answered all questions in random order in a formative paper-and-pencil test that was offered leading up to the final summative neurophysiology exam (summer 2023). For each question, students also indicated whether they thought it had been written by a human or ChatGPT.

Outcomes
The final data set consisted of 161 participants and 46 MCQs (25 human and 21 LLM questions). There was no statistically significant difference in item difficulty between the 2 question sets, but discriminatory power was statistically significantly higher in human than LLM questions (mean = .36, standard deviation [SD] = .09 vs mean = .24, SD = .14; $P = .001$). On average, students identified 57% of question sources (human or LLM) correctly.

Next Steps
Future research should replicate the study procedure in other contexts (e.g., other medical subjects, semesters, countries, and languages). In addition, the question of whether LLMs are suitable for generating different question types, such as key feature questions, should be investigated.

Problem

Test-enhanced learning is resource intensive

Numerous studies have shown that repeated testing of knowledge leads to increased retention among learners.[1] This phenomenon is called the testing effect,[2] and test-enhanced learning[3] uses this effect by providing students with repeated, ungraded tests throughout a course. In test-enhanced learning, the traditional multiple-choice question (MCQ) format is often used, as MCQs allow a reliable and valid evaluation of knowledge[4] and are a mainstay of summative exams in many medical schools the world over. However, the development of MCQs by health care professionals and medical educators is costly and resource intensive. A common rule of thumb regarding the effort involved in creating these questions is that it takes about an hour of a health care professional's or medical educator's time to develop a single high-quality MCQ. Therefore, it would be of great benefit to the training of future physicians if this process could become (at least partially) automated.

Large language models (LLMs) in medical education

The concept of the automated creation of exam questions could benefit from the recent advent of LLMs, such as ChatGPT. LLMs are systems that use natural language processing methods to "recognize, interpret, and generate text."[5(p.1930)] Following the recent hype around these artificial intelligence-based systems, which began with the release of OpenAI's ChatGPT in November 2022, a number of use cases have demonstrated how LLMs (and ChatGPT in particular) have been used to achieve results in various domains. In health care, for example, ChatGPT has been used in efforts to improve doctor–patient communication and simplify clinical management processes.

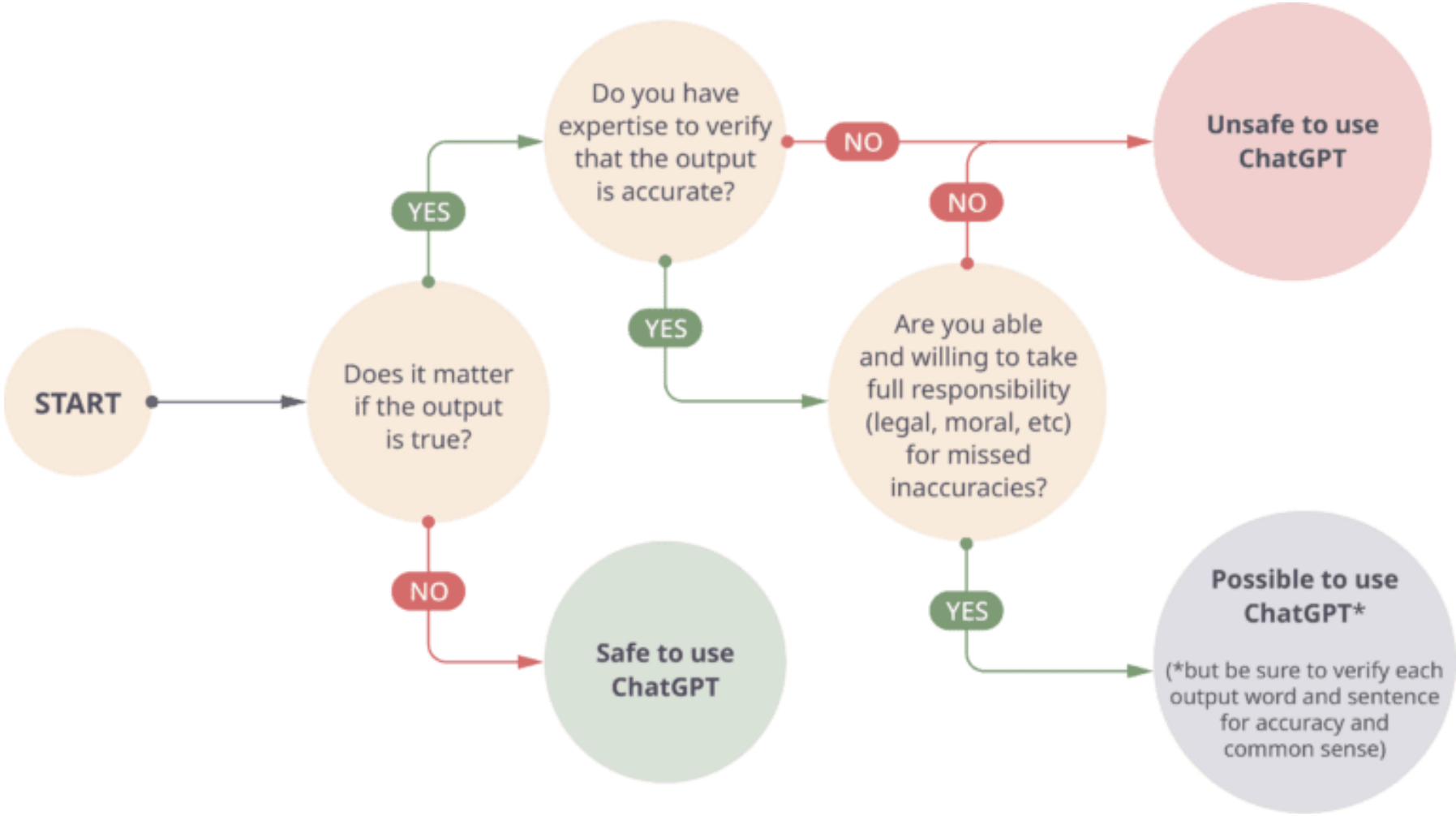
In addition to these more general applications, the advantages and disadvantages of the use of LLMs in medical education and continuing medical education have been discussed in detail. For example, Khan and colleagues describe 8 potential areas of ChatGPT application in medical education, including "teaching assistance," "personalized learning," and "creating

Please see the end of this article for information about the authors.

Correspondence should be addressed to Matthias Carl Laupichler, Institute of Medical Education, University Hospital Bonn, Venusberg-Campus 1, 53127 Bonn, Germany; telephone: +49 228 287 52167; email: matthias.laupichler@ukbonn.de.

Acad Med. 2024;99:508–512.
First published online
doi: 10.1097/ACM.0000000000005626

Responsible Use – ChatGPT & other LLMs



Professor in the Loop

- LLMs make mistakes that are hard to spot by medical students
- Output validation and/or editing by experts is required

Prompt Engineering

Dartmouth GEISEL SCHOOL OF MEDICINE | ACADEMY OF EDUCATORS AND SCHOLARS

Definition:
Crafting structured inputs to guide AI tools in generating useful and accurate outputs

Principles

- **Clarity:**
  Clearly define what you want the AI to produce

- **Context:**
  Provide sufficient background information
  (e.g., "For a group of second-year medical students…")

- **Constraints:**
  Add limits to guide format, style, or length
  (e.g., "Limit response to 200 words.")

Prompt Engineering

Open-Ended Prompts: Encourage brainstorming and diverse ideas
Example: "What are possible treatments for hypertension?"

Specific Prompts: Provide focused and precise information
Example: "Summarize the JNC 8 guidelines for hypertension management in 150 words."

Conversational Prompts: Allow for iterative back-and-forth improvement
Example: "Can you simplify this explanation for a layperson?"

Poor vs. Effective Prompts

Example 1:
Poor: "Explain diabetes.“

Effective: "Explain the pathophysiology of Type 2 Diabetes in less than 200 words for a medical student audience."

Example 2:
Poor: "Create a case study.“

Effective: "Generate a case study for a 50-year-old male presenting with chest pain, focusing on differential diagnosis and workup."

Zero-shot, Few-shot, and Multi-shot Prompting

Zero-shot Prompting

- **Definition:**
  Providing no prior examples to the AI—simply instructing it to perform a task
- **Advantages:**
  - Quick and straightforward
  - Useful for simple or general queries
- **Example:**
  Prompt: "Summarize the pathophysiology of asthma in 3 sentences."
  Output: A concise summary based on AI's training

Zero-shot, Few-shot, and Multi-shot Prompting

Multi-shot Prompting

- **Definition:**
  Providing multiple detailed examples to extensively tune the AI response on a specific task within a single prompt
- **Advantages:**
  - Produces highly accurate, tailored results
  - Best for complex or niche tasks
- **Example:**
  Prompt:
  - "Here are 3 examples of SOAP notes. Use this format to create a SOAP note for a patient with new-onset diabetes."
    (Followed by 3 fully detailed SOAP notes.)

Creating USMLE Vignette Questions

- **MCQ Bot**
  - https://tinyurl.com/GeiselVignetteBot
- Follows NBME format & guidelines
- Explanation for correct answer
- Explanation for incorrect answers

Prompt:

- You are an experienced medical educator and course director at a US medical school teaching medical students in the preclerkship phase of the MD program. One of your major goals is to prepare students for the USMLE STEP 1 exam. Create a clinical vignette exam question with NBME question writing standards. Provide an explanation for the correct answer and for each incorrect answer. Think step-by-step. Create the question based on the following topic:

RODES Prompting Framework

Dartmouth GEISEL SCHOOL OF MEDICINE | ACADEMY OF EDUCATORS AND SCHOLARS

R - Role: [Define the AI's role to set the tone and perspective of the response]

O - Objective: [Clear articulate the goal of the prompt, focusing the AI's efforts]

D - Details: [Provide specific details or parameters to guide the AI's response]

E - Examples: Here are good examples you can use to model your answer.

[Use examples to illustrate the desired style, tone, and format of the output]

S - Sense Check: Confirm the AI's understanding of the prompt, ensuring alignment before execution

RODES for Medical Education

Dartmouth GEISEL SCHOOL OF MEDICINE | ACADEMY OF EDUCATORS AND SCHOLARS

Role: You are an experienced biomedical science educator and course director at a US medical school teaching medical students.

Objective: Develop a USMLE Step 1-style question focused the following learning objective:

Relate clinical correlations to the underlying functional and anatomical organization of the somatic sensory system and describe their diagnostic value in the identification and localization of the disease processes.

Details:

- The correct answer should be: Brown-Sequard Syndrome at T10.
- Make the multiple-choice questions appropriate for 2nd year medical students preparing for STEP 1
- Include relevant patient history, physical exam findings, and any necessary laboratory results or diagnostic studies.
- The 5 answer choices (A–E) should include plausible distractors that test high-yield concepts.
- Make sure the explanation of the correct answer includes the key concepts behind both the right and wrong options.
- Think step-by-step

RODES for Medical Education

Dartmouth GEISEL SCHOOL OF MEDICINE | ACADEMY OF EDUCATORS AND SCHOLARS

Examples: Here is an examples of a good USMLE Step 1-style question:

A 67-year-old man presents to the emergency department with sudden-onset chest pain that radiates to his left arm. He is diaphoretic and pale. An ECG shows ST-segment elevations in leads II, III, and aVF. Which of the following coronary arteries is most likely occluded?

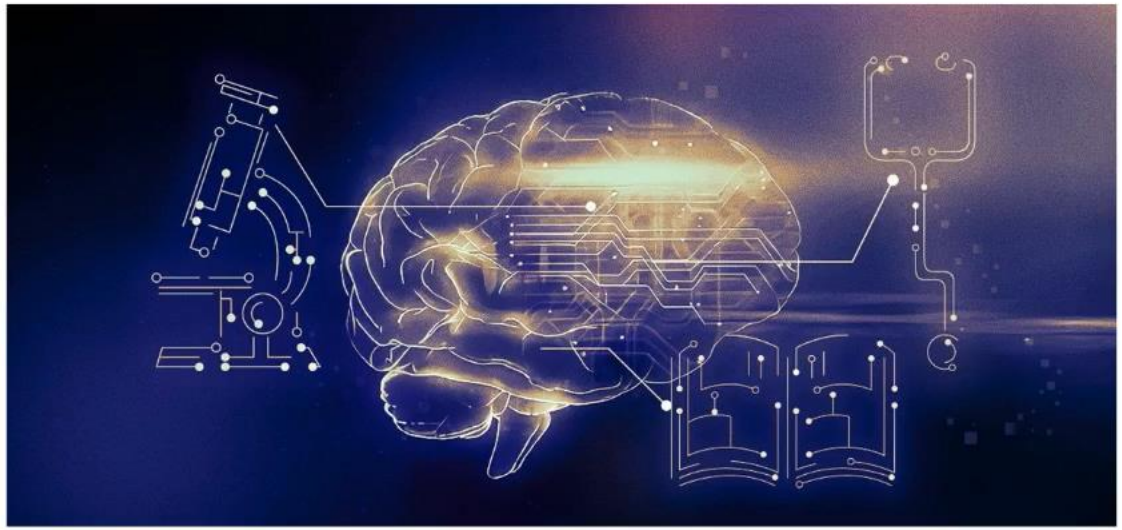
A) Left anterior descending artery
B) Left circumflex artery
C) Right coronary artery
D) Left marginal artery
E) Posterior descending artery

(Explanation: The correct answer is C. The patient's symptoms and ECG findings are consistent with an acute inferior myocardial infarction, which is most commonly due to occlusion of the right coronary artery. Distractors A and B point to other coronary vessels that are involved in different infarct locations. The explanation should explain the pathophysiology of myocardial ischemia.)

Sense Check: Do you understand the objective and the specific guidelines for creating this USMLE Step 1-style question? Do you understand the reasoning behind the correct answer?

Dartmouth GEISEL SCHOOL OF MEDICINE | ACADEMY OF EDUCATORS AND SCHOLARS

Generative AI for Medical Education Prompt-a-Thon



Thomas Thesen, Ph.D.
&
Amy Synthesia, A.I.

Group Activity # 1

Create a prompt for vignettes

1. Split into groups of 2
2. Share one laptop
3. Go to chatgpt.com
4. Scan QR code or go to tinyurl.com/MakeVignette
5. Follow instructions & work together

Case #1
Role: Medical Educator
Objective: Create a Clinical Vignette Question
Go to https://chatgpt.com/

A. Copy/Paste Prompt into ChatGPT

B. Assess Quality of Response

C. Refine Your Prompt

D. Try Something New!

#1 User Prompt

Role: You are an experienced biomedical science educator and course director at a US medical school teaching medical students.

Objective: Develop a USMLE Step 1-style question focused the following learning objective: "*Relate clinical correlations to the underlying functional and anatomical organization of the somatic sensory system and describe their diagnostic value in the identification and localization of the disease processes".*

Details:
The correct answer should be: Brown-Sequard Syndrome at T10.
•Include relevant patient history, physical exam findings, and any necessary laboratory results or diagnostic studies.
•The 5 answer choices (A–E) should include plausible distractors that test high-yield concepts.
•Keep the question at an appropriate difficulty level for 3rd year medical students.
•Make sure the explanation of the correct answer includes the key concepts behind both the right and wrong options.
•Think step-by-step.

Examples: Here is an examples of a good USMLE Step 1-style question:
A 67-year-old man presents to the emergency department with sudden-onset chest pain that radiates to his left arm. He is diaphoretic and pale. An ECG shows ST-segment elevations in leads II, III, and aVF. Which of the following coronary arteries is most likely occluded?
A) Left anterior descending artery
B) Left circumflex artery
C) Right coronary artery
D) Left marginal artery
E) Posterior descending artery
(Explanation: The correct answer is C. The patient's symptoms and ECG findings are consistent with an acute inferior myocardial infarction, which is most commonly due to occlusion of the right coronary artery. Distractors A and B point to other coronary vessels that are involved in different infarct locations. The explanation should explain the pathophysiology of myocardial ischemia.)

Sense Check: Do you understand the objective and the specific guidelines for creating this USMLE Step 1-style question? Do you understand the reasoning behind the correct answer?

Refinement suggestions:
- **Make it more relevant to your specialty**
- **Make it more relevant to your teaching at Geisel**
- **Modify the output format the way you prefer**

Group Activity # 2

Create a prompt for a clinical case

1. Split into groups of 2
2. Share one laptop
3. Go to chatgpt.com
4. Scan QR code or go to tinyurl.com/CasePrompt1
5. Follow instructions & work together

Case #2
Role: Medical Educator
Objective: Create a Clinical Case
Go to https://chatgpt.com/

A. Copy/Paste Prompt into ChatGPT

B. Assess Quality of Response

C. Refine Your Prompt

D. Try Something New!

#1 User Prompt

Role: You are a senior medical educator designing a clinical case to help preclerkship medical students understand neurology.

Objective: Create a structured clinical case presentation that encourages students to apply their knowledge of neuroanatomy and sensory pathways to localize the lesion accurately.

Details:
Use a clinical case presentation style with sections for HPI, PMH, Physical Exam, Diagnostics, and Assessment/Plan.
Ensure the case integrates high-yield neuroanatomy concepts.
Include open-ended discussion points to foster clinical reasoning.
Design the case to gradually unfold as students receive more information (mimicking real clinical presentations).
Think step-by-step

Examples: Chief Complaint:Numbness and tingling in the right hand and forearm.
History of Present Illness (HPI):A 62-year-old right-handed man presents to the neurology clinic with complaints of numbness and tingling in his right hand and forearm. The symptoms began suddenly while he was sitting at home watching television 3 days ago. He describes the numbness as "a pins-and-needles sensation" involving the thumb, index, and middle fingers, extending up the forearm. He denies any associated weakness, headaches, dizziness, or vision changes. There has been no recent trauma to the affected limb.The patient's wife confirms that there has been no slurred speech or facial droop. He is able to move his arm normally and perform daily activities without difficulty, but the numbness has remained persistent.

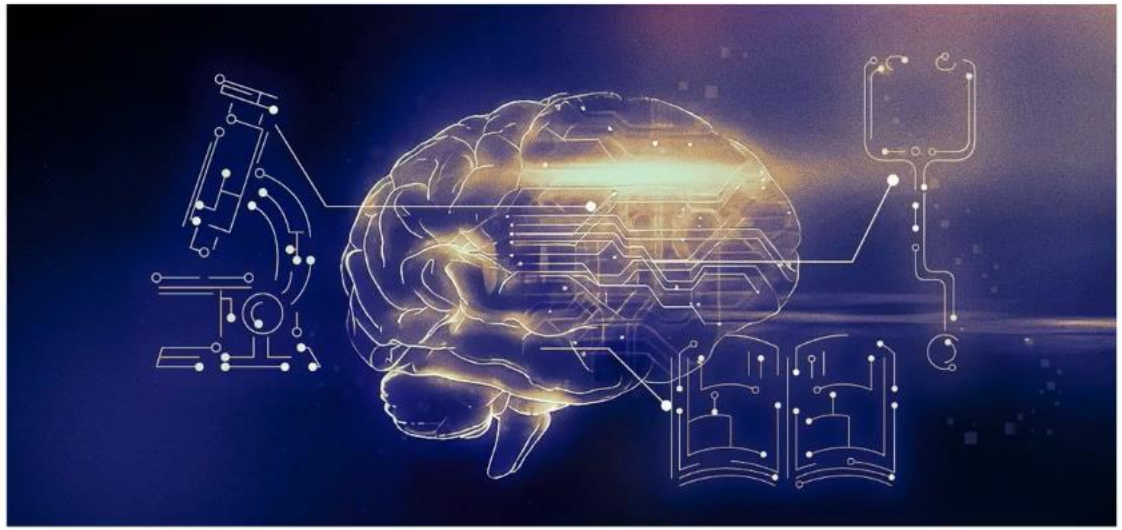
**Past Medical History (PMH):Hypertension – poorly controlledHyperlipidemiaSmoking history – 25 pack-years (quit 5 years ago)No history of diabetes or stroke
Medications:LisinoprilAtorvastatinAspirin 81 mg daily
Family History:Father had a stroke at age 65.No family history of neurological disorders. Physical Exam:General:Alert and oriented x3. No acute distress.Neurological Examination:Mental Status: NormalCranial Nerves: All intactMotor Exam:Strength: 5/5 in all muscle groups (upper and lower extremities)No pronator driftSensation:Decreased sensation to light touch, vibration, and proprioception in the right hand (thumb, index, and middle fingers) and distal forearm.Normal sensation in the face and lower extremities.Reflexes: Normal and symmetric throughout.Coordination: No ataxia or dysmetria.Gait: Normal
Diagnostics:MRI Brain:Small acute infarct in the left lateral postcentral gyrus.**

Sense Check: Do you understand the objective and the specific guidelines for creating this case?

Refinement suggestions:
- **Make prompt more relevant to your specialty**
- **Make the output more detailed and extensive**
- **Modify the output format the way you prefer**

Dartmouth GEISEL SCHOOL OF MEDICINE | ACADEMY OF EDUCATORS AND SCHOLARS

Generative AI for Medical Education Prompt-a-Thon



Thomas Thesen, Ph.D.
&
Amy Synthesia, A.I.

Dartmouth | ACADEMY OF

Case #2
Role: Medical Educator
Objective: Create a student study plan & work with a knowledge source
Go to https://chatgpt.com/

A. Copy/Paste Prompts and example data

B. Assess Quality of Response

C. Refine Your Prompt

D. Try Something New!

#1 User Prompt

Given the following lecture title and associated learning objectives, create a list of common core concepts that should be studied to help with understanding the lecture content. Suggest a logical order in which to review these core concepts.

[paste the learning objectives from #2 here]

#2 Example Data – Learning Objectives - *Molecular Basis of Cancer Session*

1.Recall the definition of an oncogene and proto-oncogene, and the assays commonly used to test their oncogenic potential.
2.Describe the two mechanisms through which a retroviruses can manifest oncogenic potential.
3.Explain how growth factor receptors can be converted from proto-oncogenes to oncogenes, how this process is illustrated by the v-erb and HER2 oncogenes, and the mechanism of action of trastuzumab (Herceptin).
4.Explain how intracellular signal transducers can be converted from proto-oncogenes to oncogenes, and how this process is illustrated by the Ras/MAP kinase pathway.
5.Describe how the genetic behavior of oncogenes and tumor suppressors differ, the types of genes that are tumor suppressors.
6.Describe how cell cycle inhibitors act as tumor suppressors, and how pathogenic variants of RB and its regulators illustrate this process.
7.Describe how signaling pathways that negatively regulate growth act as tumor suppressors, and how pathogenic variants in the TGF-beta gene illustrates this process.

#3 Advanced Prompt

Given the following medical school lecture title and associated learning objectives for third-year medical students, do the following in order:

1. Create a list of common core concepts that should be studied to help with understanding the lecture content.
2. Suggest a logical order in which to review these core concepts for novices.
3. Create for each major teaching point two multiple choice questions to help medical students solidify their understanding.

[paste the learning objectives from #2 here]

Use the attached prework document as relevant context to guide your recommendations.

[attach the prework document]

Note: Free accounts can only upload one document per day]

#4 Related Session Documents

090424_Molecular Basis of Cancer Prep Notes.pdf