



Large Language Models and Pain Medicine Education: Benefits and Pitfalls for the Medical Trainee

Michael Glicksman MD¹, Sheri Wang MD², Samir Yellapragada BS BA³, Christopher Robinson MD PhD⁴, Vwaire Orhurhu MD MPH^{5,6}, Trent Emerick MD MBA⁷

¹Department of Physical Medicine and Rehabilitation, UPMC; ²Department of Anesthesiology and Perioperative Medicine, UPMC; ³University of Pittsburgh Medical School, ⁴Department of Anesthesiology, Perioperative, and Pain Medicine, Harvard Medical School, Brigham and Women’s Hospital; ⁵UPMC; ⁶MVM Health; ⁷Department of Anesthesiology and Perioperative Medicine, Chronic Pain Division, UPMC



Introduction

- AI has exponentially evolved with LLMs becoming particularly popular¹⁻²
- In medicine, these LLMs may offer enormous potential:
 - Easy to both use and access at low costs
 - Capable of quickly integrating vast amounts of open access information
 - Promote personalized learning experiences for users³
- In pain medicine, these LLMs may thus represent an exciting tool to be used in creating individualized learning experiences to teach trainees how to navigate complex clinical situations or to solidify their understandings of pain management
- Limitations of LLMs:
 - Offer the potential to ‘hallucinate’
 - May increase educational inequalities
 - May perpetuate bias, cheating, and plagiarism^{4,5}

Methods

- Approved by UPMC Quality Improvement Review Committee (#4547)
- Three of the most commonly utilized LLMs were asked a series of multiple choice questions (MCQ) to evaluate their ability to assist in learner education within pain medicine
- 17 MCQ with five answers each compiled in style of pain medicine boards
 - Topics: anatomy, neuroimaging, clinical diagnosis, patient management, medication mechanisms and complications
- LLMs were not primed with any information prior to the questions.
- As the case report is devoid of patient identifiable information, it is exempt from IRB review requirements as per UPMC policy

Table 1. LLM Correctness Based on Question Types

Question types	ChatGPT Plus	Google Bard	Bing AI
1st order (9 total)	8	4	3
2nd order (5 total)	5	3	4
3rd order (3 total)	3	2	3
All questions (17 total)	16	9	10

Conclusions

- LLMs represent a continually evolving & promising modality to assist trainees pursuing a career in pain medicine
- Limitations currently exist that may hinder the independent use of LLMs in this setting
- Training learners to engage LLMs with tailored prompts may play a pivotal role in maximizing the application of AI in medical education
- Future research exploring how LLMs may overcome these challenges is required
- Until then, LLMs may be utilized as supplementary tool within pain medicine trainee education and with caution

Table 2. Performance of LLMs on Pain Medicine MCQ

Question	ChatGPT Plus	Google Bard	Bing AI
Which of the following is NOT associated with chronic pain? A) Low socioeconomic status (SES) B) Male gender* C) Colder climate D) BMI > 30 E) Family history of chronic pain	“Among the options given, B) Male gender is NOT generally associated with chronic pain. Chronic pain is typically reported more frequently in females than in males. There are several reasons for this difference, including hormonal factors, differences in pain perception, and potential reporting biases. Here are reasons why the other options are associated with chronic pain: Low SES (A): Lower SES has been associated with a higher prevalence of chronic pain. This can be due to multiple factors, such as increased physical labor, higher levels of stress, and limited access to healthcare. Colder climate (C): Some chronic pain conditions such as fibromyalgia, osteoarthritis, and rheumatoid arthritis can worsen in colder climates. BMI > 30 (D): Obesity (BMI > 30) has been linked with chronic pain, due to increased wear and tear on the joints, chronic inflammation, and other factors. Family history of chronic pain (E): Genetic and environmental factors can play a role in chronic pain, so a family history of chronic pain may increase one’s risk.”	“The answer is C. Colder climate. Chronic pain is a complex condition that can be caused by a variety of factors, including genetics, physical injury, and psychological factors. Some of the risk factors for chronic pain include low SES, male gender, BMI > 30, and family history of chronic pain. However, there is no evidence that colder climate is a risk factor for chronic pain. In fact, some studies have shown that people who live in colder climates may actually be less likely to experience chronic pain. This is likely due to the fact that people who live in colder climates tend to be more physically active, which can help to reduce the risk of chronic pain. So the answer is (C).”	“The answer is C) Colder climate. Chronic pain is associated with low SES, female gender, BMI > 30, and family history of chronic pain.”

* = Correct answer

Results

- Results are displayed in Table 1
- 0/3 LLMs consistently referenced existing literature in their answers
- 3/3 LLMs frequently provided explanations for selected MCQ answers
- Only ChatGPT Plus consistently explained why the other MCQ answers were incorrect
- Qualitative evaluation of these LLM-provided explanations in answering 2nd- and 3rd-order questions revealed some reasoning inconsistencies
 - E.g., providing flawed information in selecting the correct answer

Discussion

- ChatGPT Plus = highest correct % of 1st-, 2nd-, and 3rd-order Q’s
- Training data: ↓ answer accuracy by Google Bard & Bing AI in 1st-order Q’s
 - Note ChatGPT Plus and Bing AI utilize similar API
- Critical thinking capabilities: LLMs performed better with higher-level Q’s
- Reasoning inconsistencies & lack of provided explanations (Table 2)
 - Only ChatGPT Plus consistently provided explanations indicating why the other MCQ answers were incorrect
 - Limited ad hoc testing: LLMs *did* provide additional answer choice explanations when prompted to
- 0/3 LLMs consistently referenced literature in their answers
 - Limited ad hoc testing: when prompted, LLMs provided citations from websites and national organizations instead of medical texts
 - Concern re. lack of LLM access to private domain information
- Study limitations: ↓ sample size of unvalidated MCQ & LLMs, potential for unintentionally training LLMs, and lack of medical trainee evaluation
- Each LLM has since undergone revisions; only reference Google Bard & Bing AI to most accurately describe the models we tested in our pilot study
 - Bing AI → Copilot (Copilot and Copilot Pro versions)
 - Google Bard → Gemini (Gemini Nano, Pro, and Ultra models)

Full Abstract:

