

# Application of artificial intelligence in the Hong Kong Medical Licensing Examination

KK Yau<sup>1</sup>, BSc, Vincent WC Ma<sup>2</sup>, BSc, Veronica YY Li<sup>3</sup>\*, MB, BS, LLM (MEL),  
Sunny CL Au<sup>4,5</sup>, MB, ChB, FHKAM (Ophthalmology)

<sup>1</sup> Faculty of Life Sciences and Medicine, School of Basic and Medical Biosciences, King's College London, London, United Kingdom

<sup>2</sup> Faculty of Life Sciences, University College London, London, United Kingdom

<sup>3</sup> Hospital Authority, Hong Kong SAR, China

<sup>4</sup> Department of Ophthalmology, Tung Wah Eastern Hospital, Hong Kong SAR, China

<sup>5</sup> Department of Ophthalmology, School of Clinical Medicine, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China

\* Corresponding author: lyy823@ha.org.hk

This article was published on 9 Apr 2026 at www.hkmj.org.

Hong Kong Med J 2026;32:Epub

<https://doi.org/10.12809/hkmj2514630>

This version may differ from the print version.

## Introduction

Artificial intelligence (AI) tools have rapidly entered the academic landscape, reshaping how students approach complex medical concepts. Given that the Hong Kong Medical Licensing Examination (HKMLE) remains a major barrier for overseas graduates seeking to practise in Hong Kong, AI tools are likely to play an increasingly important role in examination preparation. Such a shift raises key questions regarding the reliability of AI-generated responses, the associated risks and benefits, and the implications for candidate performance and clinical competence. This commentary evaluates AI as a tool for candidates' examination preparation, balancing its benefits against potential risks including misinformation and overreliance.

## Growing use of artificial intelligence in examination preparation

The appeal of AI tools is clear: they provide instant answers, interactive explanations and seemingly confident reasoning. For HKMLE candidates, who must apply their knowledge across eight medical specialties, AI offers an accessible and efficient means of obtaining information and structured guidance.

Recent studies have demonstrated that advanced large language models have achieved strong performance in medical licensing examinations worldwide. GPT-4 achieved an overall accuracy rate of 81% across various medical licensing examinations, passing 26 of 29 and outperforming the average medical student score in 13 of 17 instances.<sup>1</sup> The earliest versions of ChatGPT achieved a passing score of 60% on the United States Medical Licensing Examination without specialised training,<sup>2</sup> while GPT-4 exceeded the performance of medical students and residents on complex clinical

reasoning questions.<sup>3</sup> These findings highlight the increasing role of AI in medical education.

## Benefits

Artificial intelligence offers several advantages that support examination preparation. Many models excel at breaking down complex concepts into clear, comprehensible explanations, making them useful for candidates who study independently. Large language models have demonstrated an ability to provide personalised learning experiences through interactive tutoring and immediate feedback,<sup>4</sup> and improvement in diagnostic reasoning among medical students who utilise AI-generated questions and feedback has been demonstrated.<sup>5</sup> The capacity of large language models to outline systematic approaches to diagnostic reasoning or to summarise complex pathophysiology reinforces structured thinking. Additionally, AI tools are not limited by geographical or logistical constraints; this accessibility offers flexibility that is unmatched by conventional learning resources. Evidence suggests that approximately 90% of AI-generated responses contain at least one key element of real-world clinical understanding that may be valuable to learners.<sup>2</sup>

## Limitations

However, current AI tools have important limitations that warrant careful consideration.

## Non-deterministic responses

One challenge is variability: the same question may generate different answers depending on the phrasing or timing of the prompt, or the version of the model. This inconsistency makes it difficult to rely solely on AI-generated responses. A study revealed noticeable variability in ChatGPT's performance across repeated attempts at the same examination, with accuracy rates ranging from 74.9% to 78.0%.<sup>6</sup>

### Potential misinformation

Artificial intelligence systems may provide inaccurate or outdated content and present it with confidence, a phenomenon termed ‘AI hallucination’ or ‘AI misinformation.’<sup>7</sup> Research has shown that, in the absence of safety measures, AI chatbots ‘hallucinated’ fabricated diseases, laboratory values, and clinical signs in up to 83% of simulated cases, accepting and elaborating on false information.<sup>8</sup>

Without sufficient background knowledge, candidates may struggle to identify subtle inaccuracies in AI-generated responses. This risk is particularly concerning in clinical settings, where minor errors (eg, inappropriate clinical guidelines, incorrect dosages or inaccurate reporting of adverse effects) could compromise patient safety.<sup>9</sup> Medical students may likewise find it difficult to detect such errors because of limited clinical experience, thereby increasing the risk of misinterpretation.<sup>10</sup>

### Lack of source transparency

Most AI tools do not consistently cite sources or provide traceable references. This lack of transparency limits users’ ability to verify information against current clinical guidelines or the medical literature in a rapidly evolving field.

### Risk of overreliance

Heavy reliance on AI for explanations may discourage students from engaging in the active problem-solving necessary for the development of clinical reasoning. Research suggests that excessive use of AI can reduce cognitive engagement and increase the risks of passivity and accepting misinformation.<sup>11</sup> Such passive learning habits may impede deeper understanding and hinder a clinician’s ability to apply knowledge independently, both during the HKMLE and in future practice.

Medical education has long promoted active learning methods, such as repeated testing and retrieval practice, as these approaches demonstrate superior long-term retention compared with passive study.<sup>12</sup> The convenience of AI-generated answers may unintentionally reduce students’ engagement in the learning practices required to promote deeper thinking and the development of clinical reasoning.

## Implications for assessment of the Hong Kong Medical Licensing Examination

The HKMLE Part I aims to assess not only factual recall but also diagnostic reasoning, pattern recognition and the candidate’s ability to apply knowledge across a range of clinical contexts.<sup>13,14</sup> Overreliance on AI may lead to superficial understanding or false confidence, particularly when

inconsistent or incorrect answers are generated. Moreover, certain areas, such as ethics, genetics or rare paediatric conditions, may be more susceptible to AI-related errors. If candidates rely heavily on AI-generated explanations, they risk acquiring incomplete or inaccurate knowledge.

As AI becomes more widely used, medical educators and regulators should consider its impact on examination fairness, assessment validity and long-term clinical competence. Evidence suggests that it has become increasingly difficult to design ‘AI-proof’ examination questions, even when they are entirely novel and image-based, raising concerns about the validity of unmonitored online assessments.<sup>15</sup> These developments challenge conventional assessment structures and underscore the need to develop alternative methods that effectively evaluate clinical reasoning and judgement.

## Recommendations

To ensure the effective use of AI tools in HKMLE preparation, candidates should adopt a balanced strategy:

1. Use AI as an additional aid rather than a replacement for textbooks and clinical guidelines.
2. Verify AI responses against authoritative sources, and cross-check information across multiple references.
3. Attempt questions independently before consulting AI; this approach preserves active reasoning and diagnostic thinking.
4. Prioritise conceptual understanding rather than memorising AI-generated answers.

Medical educators may consider providing guidance to help candidates integrate AI safely into their learning routines. Educational institutions could also consider incorporate digital literacy training that includes prompt engineering, interpretation of AI output, and critical assessment of AI-generated content.<sup>11</sup> Without such guidance, students may rely excessively on AI tools rather than developing independent critical thinking skills.

## Conclusion

Artificial intelligence has shown promise in supporting HKMLE preparation by providing accessible and structured explanations. However, its limitations—including variability, occasional misinformation, limited source transparency and risk of overreliance—mean that it cannot replace conventional learning resources. When used thoughtfully, AI can enhance examination preparation; when used indiscriminately, it may hinder deeper understanding and the development of essential clinical reasoning skills.

As AI becomes increasingly integrated into medical education, its responsible use must be

encouraged to support both examination performance and future clinical practice. The medical education community must carefully balance the benefits of technological innovation with the responsibility to train competent physicians capable of delivering safe, evidence-based patient care. Future research should examine the direct impact of AI-assisted learning on clinical performance and patient outcomes, with the aim of determining how such tools can be optimally integrated into medical training.

### Author contributions

Concept or design: SCL Au.

Acquisition of data: VWC Ma, KK Yau.

Analysis or interpretation of data: VWC Ma, KK Yau, VYY Li.

Drafting of the manuscript: VWC Ma, KK Yau, VYY Li.

Critical revision of the manuscript for important intellectual content: SCL Au.

All authors had full access to the data, contributed to the study, approved the final version for publication, and take responsibility for its accuracy and integrity.

### Conflicts of interest

All authors have disclosed no conflicts of interest.

### Funding/support

This study received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

### References

- Liu M, Okuhara T, Chang X, et al. Performance of ChatGPT across different versions in medical licensing examinations worldwide: systematic review and meta-analysis. *J Med Internet Res* 2024;26:e60807.
- Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLOS Digit Health* 2023;2:e0000198.
- Strong E, DiGiammarino A, Weng Y, et al. Chatbot vs medical student performance on free-response clinical reasoning examinations. *JAMA Intern Med* 2023;183:1028-30.
- Thesen T, Park SH. A generative AI teaching assistant for personalized learning in medical education. *NPJ Digit Med* 2025;8:627.
- Kıyak YS, Emekli E, İş Kara T, Coşkun Ö, Budakoğlu İİ. AI teaches surgical diagnostic reasoning to medical students: evidence from an experiment using a fully automated, low-cost feedback system. *J Surg Educ* 2025;82:103639.
- Lai UH, Wu KS, Hsu TY, Kan JK. Evaluating the performance of ChatGPT-4 on the United Kingdom Medical Licensing Assessment. *Front Med (Lausanne)* 2023;10:1240915.
- Hatem R, Simmons B, Thornton JE. A call to address AI “hallucinations” and how healthcare professionals can mitigate their risks. *Cureus* 2023;15:e44720.
- Littrell A. AI chatbots lack skepticism, repeat and expand on user-fed medical misinformation. *Medical Economics*. 2025 August 7. Available from: <https://www.medicaleconomics.com/view/ai-chatbots-lack-skepticism-repeat-and-expand-on-user-fed-medical-misinformation>. Accessed 30 Mar 2026.
- Tenajas R, Miraut D. The hidden risk of AI hallucinations in medical practice. *Ann Fam Med*. 2025. Available from: <https://www.annfammed.org/content/hidden-risk-ai-hallucinations-medical-practice>. Accessed 30 Mar 2026.
- Kim Y, Jeong H, Chen S, et al. Medical hallucination in foundation models and their impact on healthcare. *medRxiv*. 2025 March 3. Available from: <https://www.medrxiv.org/content/10.1101/2025.02.28.25323115v1>. Accessed 30 Mar 2026.
- Izquierdo-Condoy JS, Arias-Intriago M, Tello-De-la-Torre A, Busch F, Ortiz-Prado E. Generative artificial intelligence in medical education: enhancing critical thinking or undermining cognitive autonomy? *J Med Internet Res* 2025;27:e76340.
- Arango-Ibanez JP, Posso-Nuñez JA, Díaz-Solórzano JP, Cruz-Suárez G. Evidence-based learning strategies in medicine using AI. *JMIR Med Educ* 2024;10:e54507.
- The Medical Council of Hong Kong. Licensing Examination Information Portal. Sample Questions of Part 1 Examination. March 2021. Available from: <https://leip.mchk.org.hk/EN/print/SampleQuestionsP1-EN.pdf>. Accessed 15 Sep 2025.
- The Medical Council of Hong Kong. Licensing Examination Information Portal. Licensing Examination. Available from: [https://leip.mchk.org.hk/EN/dexam\\_I.html](https://leip.mchk.org.hk/EN/dexam_I.html). Accessed 26 Mar 2026.
- Newton PM, Summers CJ, Zaheer U, et al. Can ChatGPT-4o really pass medical science exams? A pragmatic analysis using novel questions. *Med Sci Educ* 2025;35:721-9.