OSCEai dermatology: augmenting dermatologic medical education with Large Language Model GPT-4

OSCEai dermatologie : enrichir l'enseignement médical en dermatologie grâce au grand modèle de langage GPT-4

Ye-Jean Park,[1] Eddie Guo,[2,3] Muskaan Sachdeva,[1,4] Bryan Ma,[2,5] Sara Mirali,[1,4] Brian Rankin,[2,5] Nikki Nathanielsz,[1,4] Abrahim Abduelmula,[1,4] Tatiana Lapa,[1,4] Mehul Gupta,[2,6] Trevor Champagne[1]

[1]Division of Medicine, Department of Medicine, Temerty Faculty of Medicine, University of Toronto, Ontario, Canada; [2]Cumming School of Medicine, University of Calgary, Alberta, Canada; [3]Division of Neurosurgery, Temerty Faculty of Mediicne, University of Toronto, Ontario, Canada; [4]Dermatology, Department of Medicine, Temerty Faculty of Medicine, University of Toronto, Ontario, Canada; [5]Division of Dermatology, Department of Medicine, Cumming School of Medicine, University of Calgary, Alberta, Canada; [6]Division of Internal Medicine, Department of Medicine, Cumming School of Medicine, University of Calgary, Alberta, Canada

Correspondence to: Ye-Jean Park; Medical Sciences Building, 1 King's College Cir, Toronto, ON M5S 1A8 Toronto, Ontario, Canada; email: yejean.park@mail.utoronto.ca

Implication Statement

OSCEai Dermatology demonstrates how large language models (LLMs) like GPT-4 can be integrated into medical education to enhance trainees' history taking and management skills in an OSCE-like format, including in visual-based specialties like dermatology. By generating diverse, realistic skin cancer role-play scenarios across different skin tones alongside the integration of pre-existing, evidence-based images, the app provides learners with valuable, personalized feedback. This innovation offers a novel, interactive learning tool that supplements traditional teaching methods and can be applied across various specialties. Institutions can adopt or adapt similar LLM-driven educational tools to introduce trainees to a wider range of clinical cases, fostering improved diagnostic skills and patient-centred, culturally sensitive care.

Énoncé des implications de la recherche

La dermatologie illustre comment les grands modèles de langage (GML), tels que GPT-4, peuvent être intégrés à la formation médicale pour renforcer les compétences des stagiaires en anamnèse et en prise en charge dans un format de type ECOS, y compris dans des disciplines visuelles comme la dermatologie. En générant des scénarios de jeu de rôle variés et réalistes sur le cancer de la peau, adaptés à différents tons de peau, et en intégrant des images probantes déjà existantes, l'application offre aux apprenants une rétroaction personnalisée et pertinente. Cette innovation propose un outil d'apprentissage interactif et novateur qui complète les méthodes d'enseignement traditionnelles et peut être appliqué à diverses spécialités. Les établissements peuvent adopter ou adapter des outils pédagogiques similaires fondés sur les GML afin d'exposer les stagiaires à une plus grande diversité de cas cliniques, favorisant ainsi le développement de compétences diagnostiques et de soins centrés sur le patient et culturellement sensibles.

Introduction

Large language models (LLMs), including OpenAI's ChatGPT, are emerging as powerful systems capable of processing a large corpus of data and answering free-text queries with natural language responses.[1] Despite their potential to revolutionize medical care, the application of LLMs in medical education across various specialties and the perspectives of the trainees who would be utilizing such technologies remain largely unexplored. By integrating LLMs in OSCE practice resources, learners could continue enhancing their clinical skills to diagnose rare and common conditions in dermatology.

Description of the innovation

Diagnosing skin cancer remains a significant challenge due to its diverse presentations and common mimickers.[2] Furthermore, certain skin cancers are rising in prevalence, causing significant morbidity and mortality.[2] Given this, we developed a novel app, OSCEai Dermatology (https://oscegpt.com/?searchGroup=Dermatology&caseDifficulty=Medium), that i) generates diverse, accurate Objective Structured Clinical Examination (OSCE) skin cancer role-play scenarios and ii) provides personalized, evidence-based feedback and SOAP notes to learners.

Utilizing GPT-4 and OpenAI's text-to-speech API, we created scenarios that span the entire Fitzpatrick skin type scale and feature varying difficulty levels (Figure 1; all original scenarios available upon request). Five dermatology residents in their PGY-1 to PGY-2 years evaluated each of our seven scenario's realism, interactivity, educational potential, and accuracy, along with the SOAP notes and feedback provided by OSCEai. Survey results consisted of a Likert scale (0-10) on realism, interactivity, and accuracy, along with anonymous written feedback from participants. This project was exempt from research ethics board review as it does not constitute research.
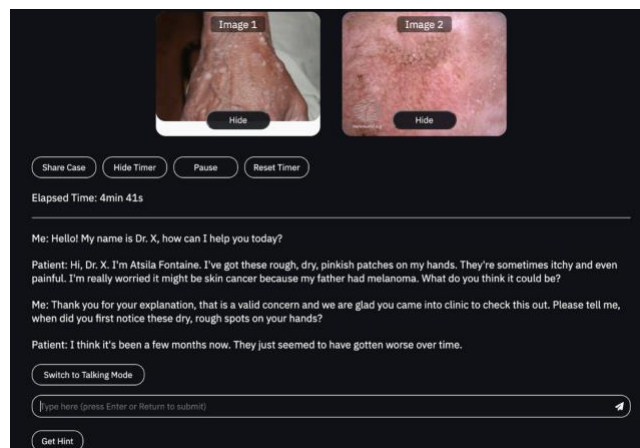
Figure 1. Scenario 1 partial example of our OSCEai Dermatology collection featuring a case of actinic keratosis. At this time, OSCEai Dermatology includes speaking/typing abilities to role-play with the patient, skin and dermoscopy images, feedback for users, and SOAP note generation features. All features available at https://oscegpt.com/.

Outcomes to date

The average scores for realism, interactivity, and accuracy of patient scenarios were 6.8 (SD: 1.29), 7.6 (SD: 2.22), and 7.4 (SD: 1.71), respectively. The accuracy of OSCEai Dermatology's feedback in role-play scenarios scored 6.4 (SD: 2.83). These scores indicate a relatively moderate level of effectiveness in preparing users for dermatology OSCE scenarios, with areas of improvement including the app's user-friendliness, the inclusion of investigation results and physical exam findings in the scenario stems, and closer alignment of the practice cases to the OSCE format (such as by asking OSCEai to provide multiple follow-up questions after a scenario). Drawbacks, however, included that the clinical images provided for patient scenarios were sometimes discrepant with the information given by OSCEai. Furthermore, there were instances of "hallucinations", where LLMs make up their own diagnosis. Qualitative feedback also included that LLM apps like OSCEai have "great potential in both refining students' approach to history-taking and testing core knowledge". Feedback and SOAP notes created by OSCEai Dermatology were well-received for their clarity and personalization. Participants further wrote that they appreciated the variety and diversity of cases, especially those featuring diverse skin tones, local versus global dermatology contexts, and adult versus pediatric dermatology

Suggestions for next steps

Limitations to our preliminary study include that our project possessed a restricted participant number across various training stages and institutions. To address this, we aim to recruit medical students, residents, and physician-educators to further evaluate perspectives on LLMs like OSCEai for medical education. Future directions to improve our app also include incorporating datasets, such as specific OSCE rubrics and instructions on how to provide culturally sensitive care. Moreover, we will continue expanding the model to cover additional specialties (e.g. RespGPT, CardioGPT).

Overall, our study demonstrated that LLMs have the potential to augment traditional teaching methods and offer dynamic learning opportunities, especially when introducing rare cases in local and global settings for patients across diverse skin colours. LLMs may ultimately empower learners in their journeys to become well-informed, capable, and culturally sensitive clinicians.

References

1. Abd-Alrazaq A, AlSaad R, Alhuwail D, et al. Large Language Models in medical education: opportunities, challenges, and future directions. *JMIR Med Educ.* 2023 Jun 1;9:e48291. http://dx.doi.org/10.2196/48291.

2. Zhang W, Zeng W, Jiang A, et al. Global, regional and national incidence, mortality and disability-adjusted life-years of skin cancers and trend analysis from 1990 to 2019: an analysis of the Global Burden of Disease Study 2019. *Cancer Med.* 2021 Jul;10(14):4905–22. http://dx.doi.org/10.1002/cam4.4046.