Ethics of Writing Personal Statements and Letters of Recommendations with Large Language Models

Logan Hostetter, Diana Kelm, and Darlene Nelson

Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, Mayo Clinic, Rochester, Minnesota

ABSTRACT

Large language models are becoming ubiquitous in the editing and generation of written content and are actively being explored for their use in medical education. The use of artificial intelligence (AI) engines to generate content in academic spaces is controversial and has been meet with swift responses and guidance from academic journals and publishers regarding the appropriate use or disclosure of use of AI engines in professional writing. To date, there is no guidance to applicants of graduate medical education programs in using AI engines to generate application content—primarily personal statements and letters of recommendation. In this Perspective, we review perceptions of using AI to generate application content, considerations for the impact of AI in holistic application review, ethical challenges regarding plagiarism, and AI text classifiers. Finally, included are recommendations to the graduate medical education community to provide guidance on use of AI engines in applications to maintain the integrity of the application process in graduate medical education.

Keywords:
graduate medical education; artificial intelligence

Medical education is rigorous, requiring empathy and integrity to develop leaders, educators, and compassionate clinicians to serve our communities. Graduate medical education (GME) and the admission process is designed to select individuals who demonstrate these capabilities and represent diverse backgrounds. Currently,

(Received in original form April 15, 2024; accepted in final form June 17, 2024)

Disclaimer: Artificial intelligence engines were not used to generate this article. ChatGPT was used to create sample letters of recommendation and personal statements not included in this article.

Correspondence and requests for reprints should be addressed to Darlene Nelson, M.D., M.H.P.E., Division of Pulmonary and Critical Care Medicine, Gonda 18 South, Mayo Clinic, 200 First Street S.W., Rochester, MN 55905. E-mail: nelson.darlene@mayo.edu.

ATS Scholar Vol 5, Iss 4, pp 486–491, 2024

DOI: 10.34197/ats-scholar.2024-0038PS

the personal statement and letters of recommendation offer a unique perspective into GME program applicants not easily reflected in other parts of the application. Others have published on the importance of the personal statement showing that most pulmonary critical care program directors cite the personal statement as at least "somewhat important" in the application process (1).

Large language models (LLMs) represent artificial intelligence (AI) trained on datasets of human content to generate human-like responses based on prompts. Several open access and paid versions of LLMs are available, including, for example, Llama from Meta; Bard or Gemini Pro from Google; Claude.ai or Mistral from Anthropic; Copilot from Microsoft; and, frequently highlighted in the press, GPT 3.5 or 4 (ChatGPT). These LLMs introduced over the last few years are already being explored to supplement medical education (2–5). Multiple entities are already implementing calls to action addressing ethics, equity, liability, and delivery of AI use in health care (4–6). With AI becoming more pervasive in medical education, it is important to consider how indistinguishable AI-generated content may be used by applicants to generate personal statements and letters of recommendation for GME training programs. This perspective reviews how AI engines may impact written components of GME applications, including perceptions of AI in professional writing, ethical principles, and AI text classifiers. Finally, we provide recommendations for the GME community to create guiding principles regarding the generation of application content with AI engines.

PERCEPTIONS OF AI IN PROFESSIONAL WRITING

The personal statement used in residency and fellowship applications offers a unique component to holistic application review. In a survey of program directors of pulmonary and critical care programs, program directors believed the personal statement to reflect attention to detail, ability to communicate, career goals, resilience, humility, and character outside of medicine, potentially impacting interview offers (1). AI may generate a well-read personal statement; however, the intent to express an applicant's character may be diluted or perceived as disingenuous. A few studies have already suggested this with readers' inability to distinguish AI-generated material from that from human applicants (7). Personal statements subsequently identified to be generated by LLMs (correctly or not) may then reflect negatively on the applicant.

Similar authenticity concerns should be considered when writing letters of recommendation. It may be tempting to use AI-generated praise for a candidate when an author is otherwise having difficulty writing a recommendation because using AI may provide a feeling of distancing the author from the recommendation. AI-written letters may then have the unintended consequence of reflecting poorly on the person being recommended if the document is interpreted as AI.

HOLISTIC APPLICATION REVIEW

Socioeconomic disparities should also be considered, and use of AI may disproportionately affect those underrepresented in medicine. There is likely unequal access to paid versions of LLMs, which may decrease the effectiveness of the users' ability to generate AI-driven content. User effectiveness in generating a "prompt" to determine the result is also likely to generate bias against those from lower socioeconomic backgrounds or those who speak

English as a second language, owing to significant output difference depending on the prompt (8). A person with early education using AI engines is more likely to generate guiding prompts to produce an effective product. Consequently, an applicant with significant exposure to prompt generation may be able to craft a well-read personal statement that is more difficult to distinguish from AI. Conversely, someone with limited exposure might generate easily identifiable AI content and thus may receive unequitable scrutiny of the overall application. Current publication of AI-generated work has already been shown to demonstrate biases against low-income communities (9). Also, consider that AI may generate sex- and/or race-based biased language already perpetuated in medical education because AI is trained on existing content primarily produced in high-income English-speaking countries (9–11).

Increasingly, GME programs are moving toward efficient processing of large volumes of applications. The Association of American Medical Colleges (AAMC), which operates the centralized Electronic Residency Application Service (ERAS), has partnered with external collaborators, offering the promise of efficient application processing as well as holistic and blinded application review. Although there is an overwhelming need for holistic and unbiased application review, these tools may be less useful if applied to AI content. ERAS partner Thalamus offers a fee-for-service platform called Cortex. As advertised, the software uses natural language processing (an AI-driven text analysis) to "promote holistic application review" and process content from application material to identify keywords (12). Language processing of applications may then have the unintended consequence of applicants who were better prepared in prompt generation to hit on keywords and subsequently score higher by natural language processing programs.

PLAGIARISM AND AUTHORSHIP

Authenticity and credibility are important in medical writing, and the use of LLMs may lead to plagiarism and misrepresentation. AI has been known to generate convincing references—known as a "hallucination citation"—or may plagiarize content without appropriate citation (13). In academic journals, some authors have begun to credit AI as a coauthor. This has prompted medical journals to generate guiding statements regarding authorship criteria (14). Other journals have stated that AI does not meet authorship criteria and should be stated in the acknowledgments because AI cannot take responsibility for its work (14). If an AI application cannot be credited as an author in scientific publications, should it be allowed to author statements about an individual's motivation and life experience, even if the user agrees with the description AI generated?

In contrast, it is well accepted and routine to use programs to review a manuscript for basic spelling and grammar or pay for professional services to review an application. AI may be an ideal editing tool, much like "spell-check," without altering the content or character of the message. LLMs could be accepted to improve spelling, grammar, and readability, particularly for applicants who speak English as a second language.

AI TEXT CLASSIFIERS

As LLMs increase in popularity, the concern for generation of professional material by LLMs has grown, leading to the introduction of AI text classifier programs.

Much like plagiarism detection tools, LLM text classifiers in theory can identify the percentage of an article generated by AI versus a human. Examples currently available include ZeroGPT, GPTZero, OpenAI, and Originality. However, these text classifiers have limitations and may erroneously label human writing as AI and vice versa (15). The present article's manuscript was analyzed with the AI detection software GPTZero and generated a classification as shown in Figure 1. (LLM was not used in the production of the manuscript.)

Text classifiers in current forms should not be used to screen all documents to avoid overclassification leading to potential inaccurate accusations and/or biases. Thus, reports should be interpreted with caution and only as an adjunct when the overall credibility of an application is in question. However, programs should reserve the possibility of using AI text classifiers for reviewing application material, because as these programs become more proficient, they could be used much like plagiarism detection software in the future. Publishing a public-facing statement that AI text classifiers may be used serves as guidance and a warning in application preparation whether the software is used or not.

GUIDANCE

AI will undoubtedly carve out a place in medical education, and it is important that training programs consider the use of AI in generation of application materials and other academic content. Review of online student networks reveals applicants seeking guidance on using LLMs in application preparation. One training program has revealed this trend and provides an excellent example of how

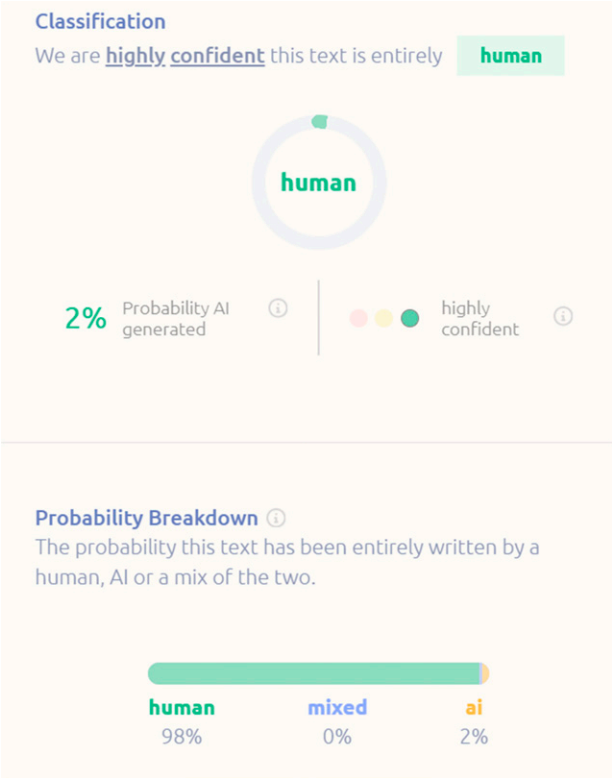

Figure 1. GPTZero program analysis of the first 5,000 characters of this article. AI = artificial intelligence.

AI engines can be used to generate a well-read personal statement in under 30 seconds (16). This evidence shows that medical organizations and training programs should provide guidance on the use of AI to generate application material and guide applicants against actions perceived as dishonest.

To date, there are no statements from large medical organizations, such as the Accreditation Council for Graduate Medical Education or the AAMC, which operates the centralized ERAS, on using LLMs to produce training program application content such as personal statements or letters of recommendation. At the time of this article's submission, the AAMC website guide on personal statements and letters of recommendation references caution regarding plagiarizing online examples of personal statements; however, there is no guidance to the use of AI engines to generate or edit application content (17). Others have suggested guidelines for use of AI in academic medicine, and these principles should be established in the application process for GME programs (18). An example of application guidance could include the following:

- If AI engines/LLMs are strictly banned from an institution's application process, this should be stated. If LLMs are used in application preparation, the following may be included:
  - AI engines/LLMs should not be directly copied to generate a letter of recommendation or personal statement.
  - AI engines/LLMs may be used for grammatical assessment and correction.
  - A disclosure statement should be included that AI engines/LLM chatbots were or were not used in preparation of text.
  - Author statement assuming integrity and accountability of the documents.
  - Statement of transparency that an LLM text classifier/plagiarism detection program may be used to review personal statements and letters of recommendation.

If we require these statements in professional publications, we should also require this guidance early in GME and generation of application content.

CONCLUSIONS

Academic journals have been quick to formulate guidance regarding authorship and AI in the academic space. As we prepare learners in GME, there is a clear need for the AAMC to provide direct guidance to applicants through ERAS. To maintain integrity and reduce applicant anxiety surrounding these issues, GME programs should be obligated to provide guidance regarding acceptable use of AI engines in application material. We hope that with this guidance, Turing testing will not become part of the application process.

Author disclosures are available with the text of this article at www.atsjournals.org.

REFERENCES

1. Hinkle L, Carlos WG, Burkart KM, McCallister J, Bosslet G. What do program directors value in personal statements? A qualitative analysis. *ATS Scholar* 2020;1:44–54.

2. Wartman SA, Combs CD. Reimagining medical education in the age of AI. *AMA J Ethics* 2019;21: E146–E152.

3. Wartman SA, Combs CD. Medical education must move from the information age to the age of artificial intelligence. *Acad Med* 2018;93:1107–1109.

4. Char DS, Shah NH, Magnus D. Implementing machine learning in health care—addressing ethical challenges. *N Engl J Med* 2018;378:981–983.

5. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. *JAMA* 2017;318:517–518.

6. Goldberg CB, Adams L, Blumenthal D, Brennan PF, Brown N, Butte AJ, *et al.*; RAISE Consortium. To do no harm - and the most good - with AI in health care. *Nat Med* 2024;30: 623–627.

7. Johnstone RE, Neely G, Sizemore DC. Artificial intelligence software can generate residency application personal statements that program directors find acceptable and difficult to distinguish from applicant compositions. *J Clin Anesth* 2023;89:111185.

8. Giray L. Prompt engineering with ChatGPT: a guide for academic writers. *Ann Biomed Eng* 2023; 51:2629–2633.

9. Curto G, Jojoa Acosta MF, Comim F, Garcia-Zapirain B. Are AI systems biased against the poor? A machine learning analysis using Word2Vec and GloVe embeddings. *AI Soc* 2022;39:617–632.

10. Koichopolos J, Ott MC, Maciver AH, Van Koughnett JAM. Gender-based differences in letters of recommendation in applications for general surgery residency programs in Canada. *Can J Surg* 2022;65:E236–E241.

11. Demzik A, Filippou P, Chew C, Deal A, Mercer E, Mahajan S, *et al.* Gender-based differences in urology residency applicant personal statements. *Urology* 2021;150:2–8.

12. Thalamus. Cortex: applicant screening re-imagined. 2022. [accessed 2024 May 8]. Available from: https://thalamusgme.com/cortex-application-screening/.

13. Athaluri SA, Manthena SV, Kesapragada VSRKM, Yarlagadda V, Dave T, Duddumpudi RTS. Exploring the boundaries of reality: investigating the phenomenon of artificial intelligence hallucination in scientific writing through ChatGPT references. *Cureus* 2023;15:e37432.

14. Stokel-Walker C. ChatGPT listed as author on research papers: many scientists disapprove. *Nature* 2023;613:620–621.

15. Rashidi HH, Fennell BD, Albahra S, Hu B, Gorbett T. The ChatGPT conundrum: human-generated scientific manuscripts misidentified as AI creations by AI text detection tool. *J Pathol Inform* 2023;14:100342.

16. Zumsteg JM, Junn C. Will ChatGPT match to your program? *Am J Phys Med Rehabil* 2023;102: 545–547.

17. Association of American Medical Colleges. Students & residents: personal statement. [accessed 2024 May 8]. Available from: https://students-residents.aamc.org/applying-residencies-eras/publication-chapters/personal-statement.

18. Kim JK, Chua M, Rickard M, Lorenzo A. ChatGPT and large language model (LLM) chatbots: the current state of acceptability and a proposal for guidelines on utilization in academic medicine. *J Pediatr Urol* 2023;19:598–604.