

Artificial Intelligence in Graduate Medical Education Applications

Sarah Mangold, DO
Margie Ream , MD, PhD

I am not a robot, and I can prove it... just by clicking on all the picture tiles with bicycles or buses. These CAPTCHA tests, or Completely Automated Public Turing tests to tell Computers and Humans Apart, protect websites from bots and spam. Will residency program directors soon need their own form of CAPTCHA to protect the integrity of the residency application process from artificial intelligence (AI)? The 2023-2024 residency application season is the first in which generative AI has been a popularly recognized and easily accessed tool. While its use has advantages, it also threatens the sincerity and reliability of important data points in application evaluation.

In November 2022, the San Francisco-based tech company OpenAI released ChatGPT—a free online chat bot capable of writing prose that is nearly indistinguishable from human-generated text. It is now one of several such services available online. Generative Pre-trained Transformer (GPT) describes large language model artificial neural networks that are “trained” on large data sets and can perform natural language processing tasks such as generating new text and analyzing text for specified content.¹ After the March 2023 update, OpenAI announced on its website:

“We’ve trained a model called ChatGPT which interacts in a conversational way. The dialogue format makes it possible for ChatGPT to answer follow-up questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests.”²

The use of AI was quickly adopted by university students. An informal survey of nearly 5000 Stanford students found that 17% used ChatGPT in fall 2022 for assignments and examinations (the term in which ChatGPT became publicly available).³ Over 25% felt that AI use, even if answers are unedited, was not a violation of honor code.³ While there are no data on AI use by health professional students, one must assume parallels with the general university population.

Because AI use is widespread at universities, many applicants will consider using it when writing a personal statement (PS). The PS has long been the applicant’s opportunity to narrate their journey through medical school and to their specialty of choice, and to highlight their unique qualities and motivations.⁴ Using AI to assist in writing them can improve efficiency, help with brainstorming, and improve communication for people for whom English is a second language.⁵ It also improves equity in access to application assistance services, which can cost in the thousands of dollars.⁶

On the other hand, multiple online forums advise against using AI for the PS because it seems “inauthentic and unoriginal—and therefore not as good as” an applicant’s own work.⁷ An AI-written PS lacks the narrative voice, context, and specific details that make it personal.⁸ Another notable drawback of AI-generated work is its propensity for lengthy and potentially excessive text, a factor that poses a barrier to reader efficiency. No data yet exist on the applicant’s perspective on the use of AI in PS generation.

Through letters of recommendation (LORs), faculty advocate for students and highlight their longitudinal relationships.⁹ Similarly to the PS, AI can assist in generating an LOR, which can improve writing efficiency and reduce the use of biased language,¹⁰ with which LORs have long been fraught.¹¹ There are no studies of how frequently AI is used to generate a professional school LOR. An article in *The Atlantic*, “The End of Recommendation Letters: Professors, Like Their Students, Use ChatGPT to Get Out of Doing Their Assignments,” reveals one popular sentiment against the use of AI in letter writing.¹² Because AI-generated letters usually lack the personal connection, emotional intelligence, nuances, and subjective judgement of a human-written letter,¹³ heavy editing is advisable if AI is used for the first draft. If AI is used in text generation, authorship is not attributed to the AI; at least this is the stance for scientific journals.¹⁴ Ultimately, the signer of an LOR takes responsibility for its content, which has always been the case.

Program directors may also look for assistance from AI. Application inflation necessitates improved

DOI: <http://dx.doi.org/10.4300/JGME-D-23-00510.1>

efficiency in application review.^{15,16} AI could streamline application screening by looking for specific combinations of content. AI could mitigate human bias, for example, by identifying biased language in an LOR,¹⁷ but could also propagate it. AI must be “trained” on the attractive features of an application, and if bias exists in the training data, that bias can be codified. When Amazon piloted the use of AI in application screening, the training data set included a majority of male employees, which unintentionally led to weighting “non-maleness” as a deficit in the application. Amazon discontinued use of this faulty paradigm once the bias was discovered.¹⁸

While AI offers advantages and threats to the application process, identifying its use is challenging. The majority of program directors reviewing the PS found AI-generated ones acceptable and did not suspect that they were not human-written.⁶ Faculty with an MD or PhD could also not consistently determine which medical research abstracts were written by humans vs AI.¹⁹ Software for AI detection is publicly available and often free, but would require additional time in the already compressed application review period, as each piece of prose would need to be uploaded or pasted into the detection site. Additionally, AI-detecting software may be inaccurate, particularly when evaluating text from a writer for whom English is a second language.²⁰ The Mozilla Foundation, whose mission is to maintain open and publicly accessible internet services, warns: “Detector tools will always be imperfect, which makes them nearly useless... One cannot accuse a student of using AI... based on the output of a detector tool that you know has a 10% chance of giving a false positive.”²¹ At least for now, it seems impossible to know with certainty if the content of prose is the product of the author and contains accurate information. However, this has always been the case. There have always been threats to authenticity and concern over the ethics of using such potentially flawed datapoints in admissions.²² AI only exacerbates them.

Currently the only guideline from the Association of American Medical Colleges on the use of AI in residency applications states: “[the PS] must be your own work and not the work of another author or the product of artificial intelligence.”²³ Applicants submitting applications to medical school through American Medical College Application Service are required to attest that “all written passages... are my own and have not been written, in part or in whole, by another author and are not the product of artificial intelligence.”²⁴ However, such an attestation is not required for residency application. These

BOX 1 ChatGPT-Generated Guidelines for Artificial Intelligence (AI) Use in Residency Applications

- 1. Respect Ethical Considerations:** Ensure that the use of AI aligns with ethical guidelines, including privacy, consent, and fairness. Avoid using AI in a way that may infringe upon others' rights or values.
- 2. Clearly Disclose AI Use:** If you choose to incorporate AI tools or algorithms into your application materials, clearly indicate which aspects were generated or reviewed using AI. Transparency is essential to maintain trust and credibility.
- 3. AI as a Supplement, Not a Replacement:** Use AI as a supplement to your application, not as a substitute for your own work and effort. Your application should primarily reflect your skills, experiences, and personal input.
- 4. Maintain Authenticity:** Your application materials should genuinely represent your qualifications, personality, and experiences. Avoid using AI to artificially enhance your qualifications or present yourself in a misleading way.
- 5. Provide Context:** Explain why you chose to use AI in your application and how it adds value to your materials. Clarify how the AI tools contributed to your presentation and understanding.
- 6. Beware of Plagiarism:** Ensure that any content generated by AI tools, such as personal statements, essays, or letters of recommendation, is unique and not plagiarized. Plagiarism is a serious offense in academic and professional settings.
- 7. Check for Errors:** AI-generated content may have errors or inaccuracies. Always review and proofread AI-generated materials to correct any mistakes and ensure the highest quality.
- 8. Be Prepared to Discuss AI Use:** If invited for an interview or during the selection process, be ready to discuss your AI use in detail. Explain how it complements your qualifications and demonstrates your technological proficiency.
- 9. Use AI Appropriately:** Only use AI where it genuinely adds value to your application. Avoid using AI for tasks that are traditionally expected to be completed by the applicant, such as personal statements or letters of intent.
- 10. Respect Program Guidelines:** Always follow the specific guidelines provided by the residency program. If the program explicitly prohibits the use of AI in application materials, respect and adhere to these rules.

These guidelines aim to strike a balance between encouraging applicants to utilize technology responsibly and ensuring that their applications reflect their genuine qualifications and abilities. The ultimate goal is to maintain fairness and integrity in the application process while allowing applicants to harness the benefits of AI when it's appropriate and ethical.

Generated in response to the instruction (generated from ChatGPT on 10/25/23): Can you produce a set of guidelines for residency programs to provide to applicants to discourage them from using AI as part of their application?

BOX 2 Author-Generated Guidelines for Artificial Intelligence (AI) Use in Residency Applications

1. **Follow and Respect Program Specific Guidelines:** Prior to submission of application materials, ensure you are aware of each program's guideline surrounding AI usage.
2. **Be Aware of Plagiarism and Maintain Authenticity:** If utilizing AI to create application materials, be aware that created elements may qualify as plagiarism. A strong application will maintain the applicant's authentic experiences and perspective.
3. **Disclose AI Use Clearly:** For all aspects of the application where AI is used, it should be clearly stated and disclosed to application reviewers.
4. **Proofread and Ensure Errors Are Corrected:** Careful review of all AI-generated materials is essential prior to submission for edits and correction of errors.
5. **Selection of AI Usage:** The applicant must carefully consider what aspects of the application could be amenable for AI usage. Avoid using AI for tasks that should reflect the individual applicant's original insight and work (personal statement, letter of intent) in order to maximize the authenticity and impact of these components of the application.

policies do not reflect the complexities of AI use. For example, is it allowable to brainstorm using AI and then edit the text heavily? Can AI be used to edit one's own work to improve grammar and readability?

Until there are more nuanced and generally agreed-upon guidelines at a national level on the use of AI in application preparation, graduate medical education (GME) programs should consider posting explicit policies on their websites regarding the use of AI in preparing application materials.²⁵ AI could even aid in the generation of such a policy (BOX 1). However, author-generated guidelines are clearer and more concise, which highlights the difference in word efficiency between AI and human text (BOX 2). Any limitations should be stated with a purpose behind them, although such policies are presently unenforceable. There have always been concerns over the use of the PS and the LOR in applicant evaluation; AI only exacerbates these and may ultimately lead to a decline in the importance of these data points in the application package. The national GME community needs to revisit the methods of holistic application review in light of the challenges posed by AI.

References

1. Hao K. What is ChatGPT? What to know about the AI chatbot. *The Wall Street Journal*. Updated May 16, 2023. Accessed April 11, 2023. <https://www.wsj.com/articles/chatgpt-ai-chatbot-app-explained-11675865177>
2. OpenAI. Accessed April 23, 2023. <https://openai.com/blog/chatgpt>
3. Cu MA, Hochman S. Scores of Stanford students used ChatGPT on final exams, survey says. *The Stanford Daily*. Published January 22, 2023. Accessed March 22, 2024. <https://stanforddaily.com/2023/01/22/scores-of-stanford-students-used-chatgpt-on-final-exams-survey-suggests/>
4. Moulton M, Lappe K, Raaum SE, Milne CK, Chow CJ. Making the personal statement "truly personal": recommendations from a qualitative case study of internal medicine program and associate program directors. *J Grad Med Educ*. 2022;14(2):210-217. doi:10.4300/JGME-D-21-00849.1
5. Woodfin MW. The personal statement in the age of artificial intelligence. *Acad Med*. 2023;98(8):869. doi:10.1097/ACM.0000000000005266
6. Johnstone RE, Vallejo MC, Zakowski M. Improving residency applicant personal statements by decreasing hired contractor involvement. *J Grad Med Educ*. 2022;14(5):526-528. doi:10.4300/JGME-D-22-00226.1
7. Whitford E. A computer can now write your college essay—maybe better than you can. *Forbes*. Published December 9, 2022. Accessed April 23, 2023. <https://www.forbes.com/sites/emmahwhitford/2022/12/09/a-computer-can-now-write-your-college-essay—maybe-better-than-you-can/>
8. Rim C. ChatGPT's college admissions essay made these 5 common mistakes. *Forbes*. Published April 5, 2023. Accessed April 23, 2023. <https://www.forbes.com/sites/christopherrim/2023/04/05/chatgpts-college-admissions-essay-made-these-5-common-mistakes/?sh=4f7b7cd575cb>
9. Chopra D, Joneja M, Sandhu G, Smith CA, Spagnuolo CM, Hookey L. Reference letters for subspecialty medicine residency positions: are they valuable for decision-making? Results from a Canadian study. *BMC Med Educ*. 2020;20(1):350. doi:10.1186/s12909-020-02270-7
10. Leung TI, Sagar A, Shroff S, Henry TL. Can AI mitigate bias in writing letters of recommendation? *JMIR Med Educ*. 2023;9:e51494. doi:10.2196/51494
11. Grimm LJ, Redmond RA, Campbell JC, Rosette AS. Gender and racial bias in radiology residency letters of recommendation. *J Am Coll Radiol*. 2020;17(1 Pt A):64-71. doi:10.1016/j.jacr.2019.08.008
12. Bogost I. The end of recommendation letters. *The Atlantic*. Published April 20, 2023. Accessed January 12, 2024. <https://www.theatlantic.com/technology/archive/2023/04/chatgpt-ai-college-professors/673796/>
13. Sreeja S. LOR written by: professionals writer Vs AI tool. *Medium*. Published July 12, 2023. Accessed October 25, 2023. <https://medium.com/@hrinfig/>

- lor-written-by-professionals-writer-vs-by-ai-tool-292b7e0b2ed#:~:text=Authenticity%20and%20Trustworthiness%3A%20Human%2Dwritten,writers%20bring%20to%20their%20recommendations
14. Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature*. 2023; 613(7945):612. doi:10.1038/d41586-023-00191-1
 15. Meyer AM, Hart AA, Keith JN. COVID-19 increased residency applications and how virtual interviews impacted applicants. *Cureus*. 2022;14(6):e26096. doi:10.7759/cureus.26096
 16. Singh NP, Boyd CJ. Rapidly increasing number and cost of residency applications in surgery. *Am Surg*. 2023;89(12): 5729-5736. doi:10.1177/00031348231173947
 17. Sarraf D, Vasiliu V, Imberman B, Lindeman B. Use of artificial intelligence for gender bias analysis in letters of recommendation for general surgery residency candidates. *Am J Surg*. 2021;222(6):1051-1059. doi:10.1016/j.amjsurg.2021.09.034
 18. Dastin J. Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. Published October 10, 2018. Accessed April 23, 2023. <https://www.reuters.com/article/idUSKCN1MK0AG/>
 19. Gao CA, Howard FM, Markov NS, et al. Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers. *NPJ Digit Med*. 2023;6(75). doi:10.1038/s41746-023-00819-6
 20. Myers A. AI-detectors biased against non-native English writers. Published May 15, 2023. Accessed March 25, 2024. <https://hai.stanford.edu/news/ai-detectors-biased-against-non-native-english-writers>
 21. Harding X. Did ChatGPT write this? Here's how to tell. *Mozilla*. Published April 14, 2023. Accessed April 23, 2023. <https://foundation.mozilla.org/en/blog/how-to-tell-chat-gpt-generated-text/>
 22. Jump J. Ethical college admissions: 'I am not a robot.' *Inside Higher Ed*. Published January 8, 2023. Accessed April 24, 2023. <https://www.insidehighered.com/admissions/views/2023/01/09/what-impact-will-chatgpt-have-college-essay-opinion>
 23. Association of American Medical Colleges. 2024 MyERAS user guide. Published 2023. Accessed October 25, 2023. <https://students-residents.aamc.org/applying-residencies-eras/publication/2024-myeras-applicant-user-guide>
 24. American Medical College Application Service. 2024 AMCAS applicant guide. Published 2023. Accessed October 25, 2023. <https://students-residents.aamc.org/media/11616/download>
 25. Zumsteg JM, Junn C. Will ChatGPT match to your program? *Am J Phys Med Rehabil*. 2023;102(6): 545-547. doi:10.1097/PHM.0000000000002238



Sarah Mangold, DO, is a PGY-3 Child Neurology Resident, Division of Child Neurology, Nationwide Children's Hospital, Columbus, Ohio, USA; and **Margie Ream, MD, PhD**, is Associate Professor, The Ohio State University College of Medicine, and Program Director, Child Neurology Residency Program, Nationwide Children's Hospital, Columbus, Ohio, USA.

Corresponding author: Margie Ream, MD, PhD, Nationwide Children's Hospital, Columbus, Ohio, USA, margie.ream@nationwidechildrens.org