doi:10.36849/JDD.8914

 BRIEF COMMUNICATION Journal of Drugs in Dermatology

Rising Influence of Artificial Intelligence: Trends in Large Language Model Usage in Dermatology Residency Personal Statements

Priya Patel Housley MS,[a] Valeria Fadhel-Hernandez BS,[b] Marjorie Montanez-Wiscovich MD PhD,[c] Dan Neal MS,[d] Mary Bohannon MHS,[c] Matthew A. Schiefer PhD[e]

[a]University of Florida College of Medicine, Gainesville, FL
[b]San Juan Bautista School of Medicine, Caguas, PR
[c]Dept of Dermatology, University of Florida College of Medicine, Gainesville, FL
[d]Department of Vascular Surgery, University of Florida College of Medicine, Gainesville, FL
[e]Malcom Randall Department of Veterans Affairs Medical Center; Dept of Biomedical Engineering, University of Florida, Gainesville, FL

ABSTRACT

With the public release of ChatGPT before the 2024 dermatology residency application cycle, applicants gained access to an advanced language model capable of generating or enhancing personal statements. This study examines trends in artificial intelligence (AI)-generated content in dermatology residency applications from 2022 to 2024 and its implications for program directors. Deidentified personal statements (n=1500) and control essays were analyzed using Quillbot, Scribbr, and ZeroGPT. A significant increase in AI-generated content was observed in the 2024 cycle compared to 2022 and 2023. Statistical tests, including ANOVA, Kruskal-Wallis, and MANOVA, confirmed year-specific differences in AI detection scores (Wilks' Lambda, $P<0.0001$). These findings highlight increasing reliance on AI tools, and future research should explore stakeholder perspectives to inform regulatory frameworks for AI usage and disclosures in residency applications.

J Drugs Dermatol. 2025;24(5):544-547. doi:10.36849/JDD.8914

INTRODUCTION

ChatGPT (OpenAI), a widely used large language model (LLM), was publicly released prior to the 2024 dermatology residency application cycle, allowing applicants access to an advanced tool capable of generating or enhancing their personal statements in a human-like manner.[1] Personal statements are one way residency programs identify applicants for interviews.[2] However, studies in specialties such as anesthesiology and plastic surgery suggest that program directors have difficulties differentiating between AI-generated and human-written essays, which raises questions about the use of LLMs in applications.[3,4] This manuscript investigates trends in LLM AI usage in personal statements to dermatology residency and explores ethical considerations of transparency in the application process.

MATERIALS AND METHODS

We deidentified and compiled dermatology personal statements submitted through ERAS during the 2022, 2023, and 2024 application cycles. We used AI to generate positive control essays using the version of ChatGPT available to applicants during their application cycle: ChatGPT-3 for 2022 and ChatGPT-3.5 for 2023 and 2024. Both models were prompted with, "Create a personal statement for someone applying to dermatology residency incorporating a person's life experiences." Personal statements known to be human-written from prior application cycles were used as negative controls.

We analyzed the applicant's personal statements (n=1500), AI-generated positive controls (n=20), and human-written negative controls (n=9) using 3 AI detection software: ZeroGPT, Quillbot, and Scribbr.[5,6,7] These tools were selected due to their reported accuracy and availability at no cost. ZeroGPT and Quillbot reported the percentage of text that was AI-generated and Scribbr reported the percent chance that the text was AI-generated.

To attain a broader understanding of the data and differences between application cycles, multiple statistical analyses were conducted in Minitab. To explore if the application cycle influenced AI detection scores across each detection software,

we conducted a one-way ANOVA followed by Tukey pairwise comparisons to identify differences between application cycles. Due to the non-normal distribution of our data, we supplemented this analysis with the Kruskal-Wallis H test. To account for the public release of ChatGPT available at no cost in November 2022, we binned the 2022 and 2023 application cycles into a "pre-availability at no cost" group to compare against the 2024 "post-availability of no cost" group.[4] Although the data were not normally distributed, given the large number of samples, we performed a two-sample t-test to compare the pre- and post-availability groups while recognizing the limitations of this statistical approach. Finally, we used MANOVA to simultaneously examine the effect of application year on AI detection across all three detection tools.

RESULTS

Our results suggest a statistically significant increase in the likelihood that personal statements contained AI-generated content in 2024 compared to previous years. During our analysis, sensitivity was defined as the ability of an AI detection tool to correctly identify known AI-generated content, while specificity was defined as the ability of the tool to correctly identify known human-written content. Accuracy measures the overall effectiveness of the tool in distinguishing control AI-generated text from known human-written text. To evaluate these metrics, we analyzed our positive and negative controls. The AI detection tools consistently displayed a high sensitivity of at least 99.2% ± 2.4% and a specificity of at least 94.49% ±11.3. Quillbot demonstrated the highest sensitivity and accuracy of 99.5 % ±2.3 and 98.4% ±5.3, respectively. Both Quillbot and Scribbr displayed the highest specificity of 97.22% ±8.33. Amongst the three tools, the lowest values of sensitivity, specificity, and accuracy calculated were 99.2% ± 2.4 (Scribbr), 94.49% ±11.3 (ZeroGPT), and 96.9% ± 6.8 (ZeroGPT), respectively. Fisher's exact tests verified that all three software were able to accurately distinguish positive and negative controls (P<0.001).

For applicant personal statements, mean differences in AI detection between each year within a 95% confidence interval is represented in Figure 1. Quillbot reported the percentage of text that was AI generated as 11.6% (Q1 = 0%, Q3 = 19.2%), 14.4% (Q1 = 0%, Q3 = 29%), and 37.9% (Q1 = 0%, Q3 = 66%) for 2022, 2023, and 2024, respectively. Scribbr showed the percent chance that the applicant personal statement was AI generated as 13.5% (Q1 = 12%, Q3 = 18%), 14.3% (Q1 = 12%, Q3 = 19%), and 30.3% (Q1 = 19%, Q3 = 46.8%) for 2022, 2023, and 2024, respectively. ZeroGPT reported the percentage of text generated by AI as 3.6% (Q1 = 0%, Q3 = 4.4%), 3.3% (Q1 = 0%, Q3 = 4.5%), and 4.6% (Q1 = 0%, Q3 = 5.8%) for years 2022, 2023, and 2024, respectively.

ANOVA results indicated a statistically significant difference in AI detection across the three application years for Quillbot (F (2,1497) = 157.50, P<0.0001), Scribbr (F (2,1497) = 136.76, P<0.0001), and ZeroGPT (F (2,1497) = 6.39, P=0.002). Additionally, with all three detectors, Tukey's post-hoc testing suggests that in 2024 AI detection was significantly different than in 2023 and 2022 while no such difference was found between 2023 and 2022 using a 95% confidence interval. A Kruskal-Wallis H Test was conducted to account for potential non-normality of the data and was found to further support statistically significant differences across all years for Quillbot (H = 216.93, P<0.0001) and Scribbr (H = 110.65, P<0.0001). For ZeroGPT, test results were marginally below the threshold of statistical significance (P=0.067).

FIGURE 1. Mean differences in AI detection percentages between application cycles (95% CI). Mean differences in AI detection percentages between application cycles for Quillbot, Scribbr, and ZeroGPT. Application cycle year comparisons are 2023 vs 2022, 2024 vs 2022, and 2024 vs 2023. Quillbot, Scribbr, and ZeroGPT indicated significant increases in AI detection from 2022 to 2024 and 2023 to 2024. 95% confidence intervals are shown.

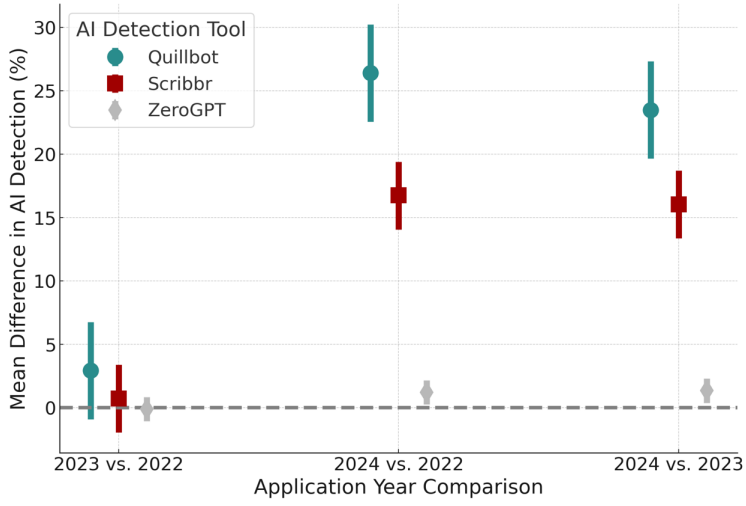
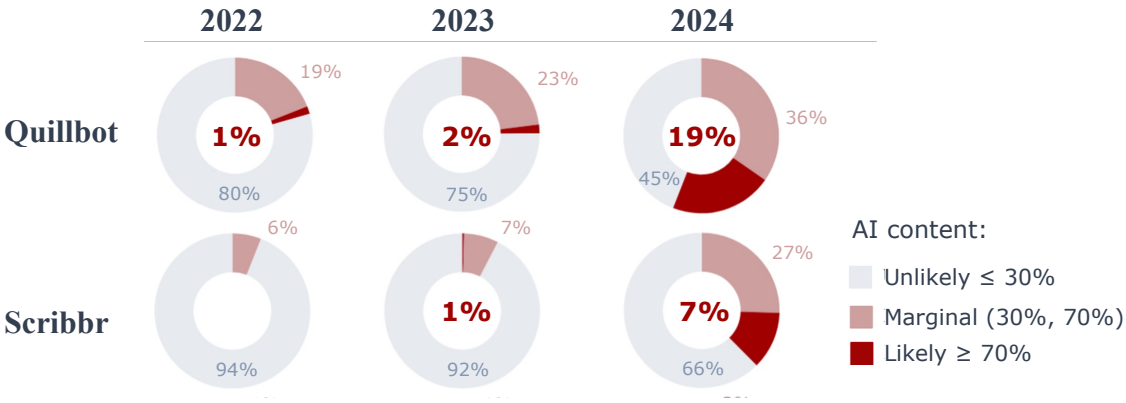


FIGURE 2. Yearly distribution of AI detection likelihood with Quillbot and Scribbr. Each chart represents the percentage of essays classified as unlikely, marginally likely, or likely to have been AI generated per year and per tool.



When comparing personal statements for application cycles before ChatGPT was available at no cost (2022 and 2023) to the cycle after its public release at no cost (2024), t-tests revealed a statistically significant increase in AI detection across all three tools ($P \leq 0.001$; Quillbot t(694) = -15.10; Scribbr t(570) = -12.79; ZeroGPT t(802) = 3.27).

Finally, a MANOVA was conducted to examine the overall effect of application year on AI detection across all three tools and indicated a statistically significant effect of year on AI detection throughout all multivariate test criteria: Wilks' Lambda ($\Lambda = 0.80$, F(6, 2990) = 60.17, $P < 0.0001$), Lawley-Hotelling (Trace = 0.26, F(6, 2988) = 63.67, $P < 0.0001$), and Pillai'sTrace (V = 0.20, F(6, 2992) = 56.67, $P<0.0001$). These findings suggest that the application year had a significant influence on AI detection scores across all detection tools.

The results of AI detection broken down by year and detection tool are further illustrated in Figure 2 which categorizes the percentage of essays as unlikely (≤ 30%), marginally likely (30%, 70%), and likely (≥ 70%) to have AI usage. This data reveals a higher percentage of essays classified as likely having AI involvement in 2024 with Quillbot and Scribbr, compared to previous application cycles and complements the statistical findings discussed previously. For transparency, although ZeroGPT suggested a significant increase in AI detection, it was excluded from Figure 2 as the reported scores did not fall within our criteria for "likely." This discrepancy reflects a difference in the tool's method of scoring text.

DISCUSSION

This study aimed to uncover trends in AI content usage within dermatology personal statements across the 2022, 2023, and 2024 application cycles by leveraging AI detection tools (Quillbot, Scribbr, ZeroGPT). Our findings revealed a statistically significant increase in detection of AI-generated content in 2024 compared to 2022 and 2023, as measured by all three detectors. These findings suggest a rise in the number of applicants who utilized LLMs to generate or enhance their dermatology personal statements. This result could indicate applicants' growing reliance on LLMs, the implications of which are discussed below. These findings are consistent with previous literature that examined higher education students' perspectives of AI usage in essays. A survey of 245 undergraduate students found that the majority of participants supported the idea of employing AI as an assistant to writing, with 44.5% "strongly agree" and 34.7% "agree" that AI should be used as a writing helper.[8]

The observations of a marked increase in AI usage in dermatology personal statements are of relevance to residency program directors who may use these essays to evaluate candidates' personal traits and characteristics. One argument is that personal statements are meant to portray an applicant's voice, with the counterargument being that these tools may promote fairness and equality, especially for those with limited mentorship. Regardless, the prevalence of LLMs raises ethical considerations of authenticity, prompting residency programs to consider if and how AI usage should be regulated.

This study's limitations include variability in the AI detection tools; however, even though the detection tools reported values with varying meanings, all algorithms detected an increase in AI usage during the 2024 application cycle. Lastly, the human-written control essays consisted of a small sample size and may not represent the varying writing styles of applicants.

Future directions that may influence policies include gauging the perspectives of applicants and program directors. These insights may inform policy reforms and guide decisions regarding the disclosure of AI usage. Further studies may explore the broader

implications of LLM usage and examine the consequences of regulating or restricting AI usage in residency application materials.

CONCLUSION

This study found a significant increase during the 2024 application cycle compared to the previous two years, suggesting a growing reliance on LLMs. As a result, residency programs may face difficulties in discerning the authenticity of applicants' narratives. Future work entails exploring the perspectives of applicants and residency program directors regarding AI usage and examining potential regulatory frameworks.

DISCLOSURES

Dr. Marjorie Montanez-Wiscovich has no current engagements with industry. In the past year, she was a principal investigator in Incyte-sponsored clinical trials. She received an educational grant from Pfizer Global Medical Grants and educational funds from PeerView and Novartis to create hidradenitis suppurativa education materials and has served on an advisory board for Sanofi. Housley, Hernandez, Neal, Bohannon, Schiefer have no conflicts to disclose.

REFERENCES

1. OpenAI. GPT-4o. Openai.com. Available at: https://openai.com/index/hello-gpt-4o/. Accessed July 12, 2024.
2. Moulton M, Lappé K, Raaum SE, et al. 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.
3. 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. doi:10.1016/j.jclinane.2023.111185.
4. Chen J, Tao BK, Park S, et al. Can ChatGPT fool the match? Artificial intelligence personal statements for plastic surgery residency applications. *Plast Surg.* 2024;0(0). doi:10.1177/22925503241264832.
5. ZeroGPT. AI text detector. Available at: https://www.zerogpt.com/. Accessed August 1, 2024.
6. AI Detector - QuillBot. Quillbot.com. Available at: https://quillbot.com/ai-content-detector. Accessed August 1, 2024.
7. AI Detector. Scribbr. Available at: https://www.scribbr.com/ai-detector/. Accessed August 1, 2024.
8. Malik AR, Pratiwi Y, Andajani K, et al. Exploring artificial intelligence in academic essay: higher education student's perspective. *Int J Educ Res Open.* 2023;5:100296. doi:10.1016/j.ijedro.2023.100296.

AUTHOR CORRESPONDENCE

Priya Patel Housley MS
E-mail:..... Patel.priya@ufl.edu