Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Association Between AI-Associated Writing Characteristics in Program Applications and Application Outcomes: A Retrospective Observational Study

Siona Samanta

samanta.siona@gmail.com

University of California, Davis Health

Robert A. Nelson

University of California, Davis Health

Neha Ramesh

University of California, Davis Health

Sophia H. Aclan

University of California, Davis Health

Simranpreet K. Bains

University of California, Davis Health

Peter S. Gorge

University of California, Davis Health

Patrick H. Hidalgo

University of California, Davis Health

Donovan S. Nielsen

University of California, Davis Health

Gabrielle J. Osorio

University of California, Davis Health

Abinaya Senthil

University of California, Davis Health

Samiksha M. Shah

University of California, Davis Health

Pranav P. Shetty

University of California, Davis Health

William T. Tuter

University of California, Davis Health

Daniel K. Nishijima

University of California, Davis Health

Research Article

Keywords: AI text detection, applicant selection, generative artificial intelligence, medical education, student writing, program applications

Posted Date: November 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11098848/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Use of artificial intelligence (AI) tools in program applications is increasing, yet their impact on application review and selection remains unclear. We examined the association between AI use, reviewer scores, and selection outcomes in a competitive emergency medicine undergraduate research associate program.

Methods

We conducted an observational study of applications submitted to an emergency medicine undergraduate research associate program from 2022–2025. Applications consisted of 4 questions (<500 words per response) and were scored by 2 independent reviewers (range 0-100). We estimated AI use on responses using a commercial AI-detection software (Winston AI) with higher scores (range 0-100) indicating greater likelihood of AI use. Outcomes included overall application score, interview invitation, and program acceptance. We conducted analyses evaluating the likelihood of AI use in predicting outcomes adjusting for application year, quarter, gender, race/ethnicity, class year, and word count. We also conducted a sensitivity analysis evaluating the association between likelihood of AI use and acceptance restricted to applicants who received interviews.

Results

Among 1,801 applications, 30% received interview invitations and 8% were accepted. Likelihood of AI use increased each year. Higher likelihood of AI use was independently associated with higher scores (p<0.001). Applications with the highest likelihood of AI use (90-100%) were also associated with interview invitations (adjusted odds ratio [aOR] 4.1, 95% CI 2.9–5.7) and acceptance (aOR 1.8, 95% CI 1.0-3.0). When restricted to applicants receiving interviews, likelihood of AI use was not independently associated with acceptance.

Conclusion

Likelihood of AI use in program applications increased each year and was associated with higher reviewer scores and interview invitations but was not independently associated with program acceptance among interviewed applicants. These findings highlight emerging challenges in evaluating written application materials and the need for clearer guidance regarding AI use in program applications.

Background

The rapid adoption of generative artificial intelligence (GAI) tools has transformed how individuals create and edit written content.[1] Large language models such as ChatGPT and other artificial intelligence (AI)-assisted writing platforms are increasingly used in academic, professional and personal settings, raising questions regarding the authenticity, evaluation, and interpretation of written work.[2–4] Many educational, research, and professional programs utilize written application materials to assess applicants'

communication skills, motivation, lived experiences, and fit for a given role.[5] The emergence of generative AI has challenged traditional assumptions underlying these evaluation processes by influencing how applicants generate and revise their written responses. Consequently, questions have emerged regarding whether AI-assisted writing influences reviewer perceptions and selection outcomes, and whether essay-based application processes remain effective measures of applicant qualifications.[6]

Existing literature and discussions have largely focused on the ethical considerations, detection strategies, and theoretical implications of generative AI usage.[1,7] However, as interest in the real-world applications and impacts of GAI continues to grow, a gap exists in the literature regarding its role in competitive application processes. It remains unclear whether applications that exhibit characteristics suggestive of AI assistance are evaluated differently by reviewers or whether such applications are associated with increased success during applicant selection. Understanding these relationships is critical for programs that rely on written responses to assess applicant qualifications and make selection decisions.

The Emergency Medicine Research Associate Program (EMRAP) is a competitive undergraduate clinical research program based in the Emergency Department at the University of California, Davis Medical Center that uses a multi-stage admissions process consisting of written applications followed by interviews for selected candidates.[8] The increasing availability of AI-assisted writing tools during recent application cycles provides a unique opportunity to examine patterns of AI-associated writing characteristics and their association with application outcomes. Our objective was to evaluate temporal trends in AI-associated writing characteristics in EMRAP application essays and to determine whether AI confidence scores were associated with reviewer scores, interview invitations, and program acceptance. We hypothesized that indicators of AI use would increase over time and would be associated with higher reviewer-based evaluation scores and increased probability of selection into the program.

Methods

Study Design and Setting

We conducted a retrospective observational study of all the applications submitted to the Emergency Medicine Research Associate Program (EMRAP) at the University of California, Davis (UC Davis) Medical Center between Fall 2022 and Fall 2025. EMRAP utilizes a multi-stage selection process consisting of written application materials, reviewer scoring, interviews, and final acceptance decisions. Each applicant writes individual responses to four essay questions, each with a suggested word count of around 300 words, that are independently reviewed and numerically scored by two applicant reviewers from the administrative team. Interview invitations are determined solely by essay scores, whereas final acceptance decisions are based on interview performance. The UC Davis Institutional Review Board reviewed and determined it to be not human subjects research (IRB No: 2476686-1).

Data Collection and AI Assessment

We extracted application data from the EMRAP application database. Variables collected included application cycle year, application quarter, gender, race/ethnicity, class year, written application responses, reviewer scores, interview invitation status, and final acceptance status.

To evaluate the likelihood of AI-assisted writing, we analyzed each application essay using WinstonAI (Model 4.0, “Curia”), a commercially available AI text detection platform. WinstonAI generates a Human Confidence Level (HCL) score ranging from 0 to 100, representing the percentage by which the software is confident that a response was written by a human rather than generated by artificial intelligence.[9] Each written essay submitted by an applicant was assigned an HCL score. The HCL scores were averaged across all submitted written responses to generate an overall AI confidence score.

For ease of interpretation, we transformed mean HCL scores into an AI Confidence Score using the equation: **AI Confidence Score (AI Score) = 100 - HCL score**.

Higher AI scores, therefore, indicate a greater estimated likelihood of AI-assisted writing. For example, an HCL score of 30% would be transformed to an AI score of 70%, indicating that the model was 70% confident that the essay was written by AI. AI score was analyzed as both a continuous variable and a categorical variable (0 to < 10, 10 to < 50, 50 to < 90, and 90 to 100) to evaluate potential non-linear associations with application outcomes.

Outcomes

The primary outcome was the overall reviewer score, defined as the average score assigned by two independent application reviewers on a scale of 0 to 100 with higher scores indicating a stronger application. The secondary outcome included receipt of interview invitation (yes/no) and acceptance into the EMRAP program (yes/no).

Statistical Analysis

We used descriptive statistics to characterize the study cohort and evaluate temporal trends in AI confidence scores across application cycles. We used a multivariable linear regression model to evaluate for associations between applicant characteristics and our primary outcome of overall review scores. Independent variables included application cycle year, application quarter, gender, race/ethnicity, class year, average word count category, and average AI confidence score. Average essay word count was dichotomized as <290 versus ≥290 words. We used logistic regression models to evaluate factors associated with secondary outcomes of interview invitation and program acceptance. For these analyses, AI confidence scores were categorized into four groups (0 to <10, 10 to <50, 50 to <90, and 90 to 100), with the lowest category serving as the reference group. We included all eligible applications during the prespecified study period; therefore, no a priori sample-size calculation was performed.

Analyses were conducted using Stata version 14.0 (StataCorp, College Station, TX).

We performed a sensitivity analysis to determine whether associations between AI confidence scores and acceptance persisted after the interview state, by restricting the cohort to applicants who received an interview invitation. Multivariable logistic regression was then repeated using acceptance as the outcome. Adjusted mean differences, adjusted odds ratios, and corresponding 95% confidence intervals were reported. Statistical significance was defined as a two-sided p-value <0.05.

Results

Study Cohort

A total of 1,801 applications were submitted to EMRAP between the 2022 and the 2025 application cycles and analyzed in this study (Table 1). Across the four study years, 153 applications were submitted in 2022, 476 in 2023, 548 in 2024, and 624 in 2025. Applicants were predominantly female (57.02%), Asian (56.41%), and third-year undergraduate students (35.26%). The mean average word count across the four application essays was 246 (standard deviation [SD], 63) words, the median AI score was 0.2 (interquartile range [IQR], 0–50), and the mean overall reviewer score was 62.8 (SD, 22.7). Overall, 537 (30%) applicants received an interview invitation, and 151 (8%) were ultimately accepted into the program.

Table 1
Study characteristics of EMRAP applications, 2022–2025

Characteristic	n (%)
Application Cycle	
2022	153 (9)
2023	476 (26)
2024	548 (30)
2025	624 (35)
Application Quarter	
Fall	512 (28)
Winter	508 (28)
Spring	387 (21)
Summer	394 (22)
Gender	
Female	1,027 (57)
Male	418 (23)
Not reported	338 (19)
Other	18 (1)
Race/Ethnicity	
Asian	1,016 (56)
Unknown	243 (13)
Hispanic/Latino	177 (10)
Mixed	98 (5)
Middle Eastern	98 (5)
White	104 (6)
Black	51 (3)
Other	10 (1)
Class year	
Pre-college	15 (1)

Characteristic	n (%)
1st year	261 (14)
2nd year	572 (32)
3rd year	635 (35)
4th year	297 (16)
5th year	21 (1)
Word count per response, mean (SD)	246 (63)
AI score, median (IQR)	0.2 (0–50)
Overall reviewer score, mean (SD)	63 (23)
Invitation for interview	537 (30)
Acceptance into program	151 (8)

Temporal Trends in AI-Associated Writing Characteristics

The estimated use of AI increased yearly between 2022 to 2025 while average overall reviewer scores remained relatively stable (Fig. 1). Median AI scores increased from 0 in 2022 and 2023 to 29 in 2025, while the proportion of applications in the highest AI score category (90–100) increased from 0% in 2022 to 28% in 2025 (Table 2). Figure 2 demonstrates the change by estimated AI use by category over time.

Table 2
Median overall AI scores and percentage by category, by year from 2022–2025

Year, n	AI score, median (IQR) [a]	AI score 0 to < 10, n (%)	AI score 10 to < 50, n (%)	AI score 50 to < 90, n (%)	AI score 90 to 100, n (%)
2022, 153	0 (0–0)	145 (95)	8 (5)	0 (0)	0 (0)
2023, 476	0 (0–0)	405 (85)	13 (3)	18 (4)	0 (0)
2024, 548	0 (0–50)	319 (58)	90 (16)	59 (11)	80 (15)
2025, 624	29 (0–99)	242 (39)	113 (18)	95 (15)	174 (28)

[a] – Higher AI scores indicate a greater estimated likelihood of AI-associated writing.

Factors Associated with Reviewer Scores

On adjusted analysis, higher estimated AI scores were independently associated with higher reviewer scores, with reviewer scores increasing by 0.14 points for each one-point increase in AI score (95% confidence interval [CI], 0.11–0.17; $P < 0.001$) (Table 3).

Table 3
Adjusted analysis of average reviewer scores

Variable	Adjusted mean difference (95% CI)	P value
Year (Reference: 2022)		
2023	−7.31 (− 11.84, − 2.77)	0.002
2024	−11.16 (− 15.66, − 6.66)	< 0.001
2025	−11.27 (− 15.87, − 6.67)	< 0.001
Quarter (Reference: Fall)		
Winter	4.49 (1.50, 7.48)	0.003
Spring	4.60 (1.44, 7.75)	0.004
Summer	2.57 (− 0.56, 5.71)	0.108
Gender (Reference: Female)		
Male	−3.15 (− 5.62, − 0.67)	0.013
Not reported	−4.20 (− 7.45, − 0.94)	0.012
Other	8.02 (− 2.07, 18.11)	0.119
Race/Ethnicity (Reference: Asian)		
Unknown	−0.87 (− 4.57, 2.83)	0.65
Hispanic/Latino	−5.17 (− 16.61, 6.27)	0.38
Mixed	−3.10 (− 9.24, 3.04)	0.32
Middle Eastern	−1.29 (− 4.76, 2.18)	0.47
White	2.92 (− 1.58, 7.42)	0.20
Black	−0.77 (− 5.28, 3.74)	0.74
Other	−0.91 (− 5.29, 3.47)	0.68
Class year (Reference: 1st year)		
Pre-college	−18.98 (− 30.28, − 7.68)	0.001
2nd year	3.12 (− 0.07, 6.32)	0.055
3rd year	3.02 (− 0.12, 6.16)	0.059
4th year	1.67 (− 2.00, 5.33)	0.37
5th year	0.41 (− 9.29, 10.11)	0.93
Word count ≥ 290	10.31 (7.97, 12.65)	< 0.001

Variable	Adjusted mean difference (95% CI)	P value
Average AI score	0.14 (0.11, 0.17)	< 0.001

Other factors associated with reviewer scores included average word count, application cycle, application quarter, gender, and class year. Applications averaging at least 290 words received reviewer scores 10.3 points higher than shorter applications (95% CI, 7.97–12.65; P < 0.001). Compared with the 2022 application cycle, reviewer scores were significantly lower in 2023, 2024, and 2025. Winter and spring applications received significantly higher reviewer scores than fall applications, while male applicants and applications with unreported gender received lower scores than female applicants. Pre-college applicants received lower reviewer scores than first-year applicants. No significant associations were observed for race/ethnicity.

Factors Associated with Interview Invitation

On adjusted analysis, applicants with higher estimated AI scores generally had greater odds of receiving an interview invitation (Table 4). Compared with applicants in the lowest AI score category (0 to < 10), applicants with AI scores of 10 to < 50 had 1.69-fold greater odds of interview invitation (adjusted odds ratio [aOR], 1.69; 95% CI, 1.21–2.36; P = 0.002), while applicants in the highest AI score category (90 to 100) had more than fourfold greater odds (aOR, 4.06; 95% CI, 2.92–5.65; P < 0.001). Applicants with AI scores of 50 to < 90 had similar odds of interview invitation that were not statistically significant.

Table 4
Adjusted analysis of interview invitation

Variable	Adjusted Odds Ratio (95% CI)	P value
Year (Reference: 2022)		
2023	0.78 (0.49–1.24)	0.30
2024	0.42 (0.26–0.67)	< 0.001
2025	0.37 (0.23–0.60)	< 0.001
Quarter (Reference: Fall)		
Winter	1.07 (0.77–1.47)	0.69
Spring	1.48 (1.06–2.07)	0.021
Summer	0.80 (0.56–1.12)	0.20
Gender (Reference: Female)		
Male	0.85 (0.65–1.10)	0.22
Not reported	0.69 (0.48–0.98)	0.04
Other	0.69 (0.24–2.00)	0.50
Race/Ethnicity (Reference: Asian)		
Unknown	1.03 (0.69–1.55)	0.87
Hispanic/Latino	1.77 (0.55–5.70)	0.34
Mixed	1.15 (0.59–2.22)	0.68
Middle Eastern	1.07 (0.74–1.56)	0.72
White	1.50 (0.95–2.35)	0.08
Black	1.52 (0.96–2.42)	0.07
Other	1.28 (0.80–2.03)	0.30
Class year (Reference: 1st year)		
Pre-college	0.82 (0.21–3.16)	0.78
2nd year	1.32 (0.94–1.87)	0.11
3rd year	1.17 (0.83–1.64)	0.37
4th year	1.08 (0.73–1.61)	0.71
5th year	1.78 (0.67–4.74)	0.25
Word count ≥ 290	2.59 (2.05–3.29)	< 0.001

Variable	Adjusted Odds Ratio (95% CI)	P value
AI score (Reference 0 to < 10)		
10 to < 50	1.69 (1.21–2.36)	0.002
50 to < 90	1.47 (0.98–2.20)	0.07
90 to 100	4.06 (2.92–5.65)	< 0.001

Applications averaging at least 290 words had significantly greater odds of interview invitation than shorter applications (aOR, 2.59; 95% CI, 2.05–3.29; P < 0.001). Interview invitations were less likely during the 2024 and 2025 application cycles than in 2022, while spring applications had greater odds of interview invitation than fall applications. Applicants with unreported gender had lower odds of interview invitation than female applicants. No significant associations were observed for race/ethnicity or class year.

Factors Associated with Program Acceptance

On adjusted analysis, applicants in the highest AI score category had greater odds of program acceptance than applicants in the lowest AI score category (Table 5). Applicants with AI scores of 90 to 100 had 1.77-fold greater odds of acceptance than applicants with AI scores of 0 to < 10 (aOR, 1.77; 95% CI, 1.03–3.03; P = 0.037), whereas lower AI score categories were not significantly associated with acceptance.

Table 5
Adjusted analysis of program acceptance

Variable	Adjusted Odds Ratio (95% CI)	P value
Year (Reference: 2022)		
2023	1.04 (0.53–2.05)	0.92
2024	0.52 (0.25–1.06)	0.073
2025	0.52 (0.25–1.08)	0.078
Quarter (Reference: Fall)		
Winter	0.81 (0.49–1.35)	0.42
Spring	0.77 (0.44–1.33)	0.35
Summer	0.81 (0.47–1.38)	0.44
Gender (Reference: Female)		
Male	0.36 (0.18–0.74)	0.005
Not reported	1.43 (0.98–2.10)	0.065
Race/Ethnicity (Reference: Asian)		
Unknown	1.42 (0.71–2.83)	0.32
Hispanic/Latino	1.60 (0.33–7.80)	0.56
Mixed	1.29 (0.47–3.48)	0.62
Middle Eastern	1.29 (0.73–2.27)	0.38
White	0.89 (0.41–1.95)	0.78
Black	0.86 (0.40–1.84)	0.69
Other	1.08 (0.50–2.34)	0.85
Class year (Reference: 1st year)		
Pre-college	1.45 (0.17–12.24)	0.73
2nd year	1.54 (0.84–2.82)	0.16
3rd year	1.70 (0.94–3.07)	0.081
4th year	1.60 (0.81–3.17)	0.18
5th year	4.79 (1.32–17.36)	0.017
Word count ≥ 290	2.70 (1.89–3.87)	< 0.001
AI score (Reference 0 to < 10)		

Variable	Adjusted Odds Ratio (95% CI)	P value
10 to < 50	1.57 (0.94–2.62)	0.082
50 to < 90	0.90 (0.43–1.89)	0.78
90 to 100	1.77 (1.03–3.03)	0.037

Applications averaging at least 290 words also had greater odds of acceptance (aOR, 2.70; 95% CI, 1.89–3.87; P < 0.001). Male applicants had lower odds of acceptance than female applicants, while fifth-year applicants had greater odds of acceptance than first-year applicants. No significant associations were observed for application year, application quarter, or race/ethnicity.

Sensitivity Analysis Among Interviewed Applicants

Among applicants who received an interview invitation, estimated AI score was no longer associated with program acceptance after adjustment for applicant and application characteristics (Table 6).

Table 6
Adjusted analysis of program acceptance among interviewed applicants

Variable	Adjusted Odds Ratio (95% CI)	P value
Year (Reference: 2022)		
2023	1.31 (0.59–2.90)	0.504
2024	0.90 (0.38–2.10)	0.808
2025	1.11 (0.46–2.65)	0.819
Quarter (Reference: Fall)		
Winter	0.74 (0.40–1.35)	0.325
Spring	0.51 (0.27–0.97)	0.039
Summer	0.96 (0.50–1.82)	0.894
Gender (Reference: Female)		
Male	0.39 (0.18–0.84)	0.017
Not reported	1.91 (1.19–3.05)	0.007
Race/Ethnicity (Reference: Asian)		
Unknown	1.42 (0.65–3.11)	0.377
Hispanic/Latino	1.00 (0.15–6.85)	0.999
Mixed	1.37 (0.42–4.47)	0.606
Middle Eastern	1.35 (0.68–2.69)	0.389
White	0.63 (0.26–1.50)	0.299
Black	0.59 (0.25–1.41)	0.236
Other	0.86 (0.35–2.12)	0.739
Class Year (Reference: 1st year)		
Pre-college	3.18 (0.21–48.52)	0.405
2nd year	1.33 (0.68–2.60)	0.413
3rd year	1.61 (0.82–3.16)	0.164
4th year	1.49 (0.68–3.24)	0.321
5th year	6.08 (1.07–34.40)	0.041
AI score (Reference: 0 to < 10)		
10 to < 50	1.14 (0.63–2.08)	0.658

Variable	Adjusted Odds Ratio (95% CI)	P value
50 to < 90	0.67 (0.29–1.55)	0.345
90 to 100	0.65 (0.34–1.22)	0.176
Average word count ≥ 290	1.48 (0.97–2.26)	0.070

Discussion

In this retrospective analysis of applications submitted to EMRAP, AI-associated writing characteristics increased substantially following the widespread availability of generative AI tools. After adjustment for applicant and application characteristics, higher AI scores were independently associated with higher reviewer scores and greater odds of receiving an interview invitation. Although applicants in the highest AI score category had greater odds of program acceptance in the full applicant cohort, this association was no longer observed among applicants who advanced to the interview stage. These findings suggest that AI-associated writing characteristics may influence program acceptance indirectly through advancement to the interview stage rather than directly affecting final selection among interviewed applicants.

Comparison with Previous Literature

Our findings add to emerging evidence that generative AI may affect the evaluation of written materials used in competitive admissions.[10] Prior experimental studies have demonstrated that reviewers have difficulty distinguishing AI-generated from human-authored application essays and that AI-generated essays may receive evaluations comparable to or higher than human-written essays.[6, 10, 11] In our study of actual applicants, examining multiple stages of the selection process demonstrated that AI-associated writing characteristics were independently associated with reviewer scores and interview invitations, but not with final acceptance among interviewed applicants. These findings suggest that the influence of AI-assisted writing may be greatest during the initial written screening phase and attenuated when additional modes of assessment are incorporated.

Implications for EMRAP

For EMRAP, these findings have important implications for the role of written applications in the era of generative AI. The increasing prevalence of AI-associated writing characteristics suggests that the review of applicant essays may no longer assess the same constructs they were originally intended to evaluate, including motivation, lived experiences, written communication skills, and preparedness. In a clinical research program where trust, integrity, authenticity, and scientific rigor are foundational to patient-facing research and scholarship, this shift raises questions about authorship, ownership of written materials, and the appropriate role of AI assistance during the application process. As the distinction between acceptable editing support and substantive AI-generated content becomes increasingly difficult to define, programs such as EMRAP may need to reconsider what written essays are

intended to evaluate and whether application policies and reviewer expectations remain aligned with those goals.

At the same time, the absence of an association between AI score and program acceptance among interviewed applicants suggests that AI-associated writing characteristics may exert their influence earlier in the selection pathway. While written essays may affect initial reviewer impressions and advancement to interviews, final selection among interviewed applicants may depend more heavily on information obtained through additional modes of assessment, such as interviews. Interviews allow programs to evaluate qualities that cannot be delegated to AI, including interpersonal communication, professionalism, critical thinking, and the ability to discuss and reflect on personal experiences. In this way, the interview stage may help limit the influence of AI-associated writing by allowing applicants to be evaluated on qualities beyond their written materials. However, a multistage admissions process does not eliminate the potential influence of AI-associated writing on selection, because advancement to later stages remains dependent on earlier evaluations based on written materials. Even if AI-associated writing does not directly influence final selection among interviewed applicants, its association with advancement to the interview stage may indirectly affect an applicant's opportunity for eventual acceptance.

For EMRAP administrators, these findings underscore the need to reconsider whether the current application process remains the most fair and effective approach to selecting applicants in the era of generative AI. EMRAP has begun implementing a policy prohibiting AI use in application materials and has transitioned its application platform from Google Forms to Canvas, enabling potential future use of Turnitin to automatically generate AI-detection scores for submitted essays. This approach has not yet been implemented and remains under consideration because of logistical challenges associated with incorporating AI detection into application review. These changes aim to establish clearer expectations for AI use and preserve the intended role of written responses in evaluating applications. Continued evaluation will be necessary to determine whether these measures, alongside EMRAP's multistage selection process, support a fair and effective approach to application selection.

Broader Implications for Application-Based Programs

Although this study examined a single undergraduate clinical research program, the findings have broader implications for educational, research, and professional programs that rely on written application materials. Personal statements and supplemental essays have traditionally been used to assess applicants' communication skills, motivation, reflection, and organizational fit[5]. Increasing access to generative AI challenges the assumption that these materials primarily reflect an applicant's independent writing. Emerging evidence further suggests that reviewers have difficulty distinguishing AI-generated from human-written application essays and assign similar evaluations to both.[6,11] Together, these findings raise questions about whether written application materials continue to distinguish applicants based on the qualities they were originally intended to assess.

These findings should not be interpreted as evidence that AI-assisted writing is inherently beneficial or detrimental, nor do they suggest that applicants who use AI should automatically be advantaged or penalized. Rather, they highlight the need to reconsider what written application materials are intended to measure and whether current admissions processes continue to evaluate those qualities effectively. As generative AI becomes increasingly integrated into the application process, programs may need to place greater emphasis on complementary assessment methods, such as structured interviews, situational judgment tests, timed writing exercises, or other approaches that evaluate skills less susceptible to AI assistance.

Conclusion

The increasing availability of generative AI is changing how applicants prepare written application materials. Although AI-associated writing characteristics were not associated with acceptance among interviewed applicants, their association with reviewer scores and interview invitations suggests a potential indirect influence on final selection through advancement to the interview stage. Rather than focusing solely on whether applicants should use AI, programs may need to reconsider what written application materials are intended to measure and whether current admissions processes remain aligned with those goals.

Limitations

Several limitations should be considered when interpreting these findings. First, this was a retrospective observational study and therefore cannot establish a causal relationship between AI-associated writing characteristics and application outcomes. Second, the study was conducted within a single undergraduate clinical research program at one academic institution, which may limit the generalizability of the findings to other educational or professional selection processes. Third, AI scores were derived using a single commercially available AI detection platform. Although Winston AI provided a standardized method for estimating AI-associated writing characteristics, current AI detection technologies remain imperfect and continue to evolve alongside generative AI models. Accordingly, AI scores should be interpreted as estimates of AI-associated writing characteristics rather than definitive measures of AI use.

Finally, unmeasured factors may have influenced reviewer evaluations and application outcomes. Essay prompts and interview questions varied across application cycles, reviewer composition changed over time, and annual differences in applicant volume and available program positions likely affected the competitiveness of each application cycle. Although we adjusted for multiple applicant and application characteristics, residual confounding may have influenced the observed associations.

Future Directions

This study represents an initial step toward understanding the role of generative AI in competitive application processes. Future research should examine whether AI-associated writing characteristics are

associated with subsequent performance within the program. For example, investigators could evaluate research productivity, professionalism, communication skills, leadership development, retention, and supervisor evaluations to determine whether AI-associated writing characteristics reflect meaningful differences in applicant preparedness or simply differences in how applicants prepare written materials.

Additional studies are also needed to determine whether these findings generalize to other undergraduate research programs, graduate admissions, and professional hiring processes. As generative AI continues to evolve, longitudinal research will be important for understanding how applicant behavior, reviewer expectations, and selection practices change over time. Future work should also evaluate alternative approaches to written application assessment, including revised scoring rubrics, AI disclosure policies, and complementary assessment methods that preserve fairness, authenticity, and equity throughout the admissions process.

Abbreviations

AI
Artificial Intelligence
aOR
Adjusted Odds Ratio
CI
Confidence Interval
EMRAP
Emergency Medical Research Associate Program
GAI
Generative Artificial Intelligence
HCL
Human Confidence Level
IQR
Interquartile Range
SD
Standard Deviation
UC Davis
University of California, Davis

Declarations

Ethics approval and consent to participate:
The University of California, Davis Institutional Review Board reviewed the study and determined that it did not constitute human subjects research (IRB No. 2476686-1). Therefore, ethics approval and informed consent were not required.

Consent for publication:
Not applicable.

Availability of data and materials:
All data generated or analyzed during this study are included in this published article and its supplementary information file [see Additional file 1].

Competing interests:
The authors declare that they have no competing interests.

Funding:
This study received no external funding.

Authors' Contributions:
SS, SHA, PSG, DSN, GJO, AS, SMS, and DKN contributed to the study concept and design. SS, RAN, NR, SHA, SKB, PHH, GJO, AS, SMS, and WTT contributed to data acquisition. SS, DSN, and DKN analyzed and interpreted the data. DSN and DKN provided statistical expertise. SS drafted the manuscript. SS, DSN, PPS, and DKN critically revised the manuscript for important intellectual content. All authors read and approved the final manuscript.

Acknowledgements:
Not applicable.

References

1. Kasneci E, Seßler K, Küchemann S, et al. ChatGPT for good? On opportunities and challenges of large language models for education. Learn Individ Diff 2023;103:102274.
2. Pratama RD, Widiati U, Basthomi Y, Khoiri NE, Muniroh S, Omolu FA. "From textual originality to cognitive transparency": Reframing ethics and authenticity of scientific writing in the age of artificial intelligence through human-AI collaborative writing. J Acad Ethics 2026;24(3):77.
3. Ilie O-A. The ethics of AI-assisted academic writing. authenticity criteria in the evaluation of students' assignments. Int Conf Knowl Based Organ 2025;31:155-8.1.
4. The rapid ascent of AI-assisted writing: Transformations, tools, and trajectories 2022-2025. VibeCentral, 2025. (*Accessed July 5, 2026, at https://vibecentral.ai/report/writing/the-rapid-ascent-of-ai-assisted-writing-transformations-tools-and-trajectories/)*
5. de Boer T, Van Rijnsoever F, de Bresser H. Dear admission committee…: Which moves in application essays predict student master grades? PLoS One 2024;19(6):e0304394.
6. Choi J, Zhao J, Ngo T-A, Grierson L. Investigating the threat of AI to undergraduate medical school admissions: a study of its potential impact on the rating of applicant essays. Can Med Educ J

2025;16(6):15-20.

7. Eaton SE. Postplagiarism: transdisciplinary ethics and integrity in the age of artificial intelligence and neurotechnology. Int J Educ Integr 2023;19(1):23.
8. UC Davis Emergency Medicine Research Associate Program. Sacramento, CA: University of California, Davis Health. (*Accessed July 5, 2026, at https://health.ucdavis.edu/emergency/education/undergrad/emrap/index.html*)
9. Interpreting our AI detection scores. Winston AI. (*Accessed July 5, 2026, at https://gowinston.ai/interpreting-our-ai-detection-scores/*)
10. Vaccaro MJ, Sharma I, Espina-Rey AP, et al. Death of the personal statement: Qualitative comparison between human-authored and artificial intelligence-generated medical school admissions essays. J Am Coll Surg 2026;242(1):47-52.
11. Crawford LM, Hendzlik P, Lam J, et al. Digital ink and surgical dreams: Perceptions of artificial intelligence–generated essays in residency applications. J Surg Res 2024;301:504-11.

Figures

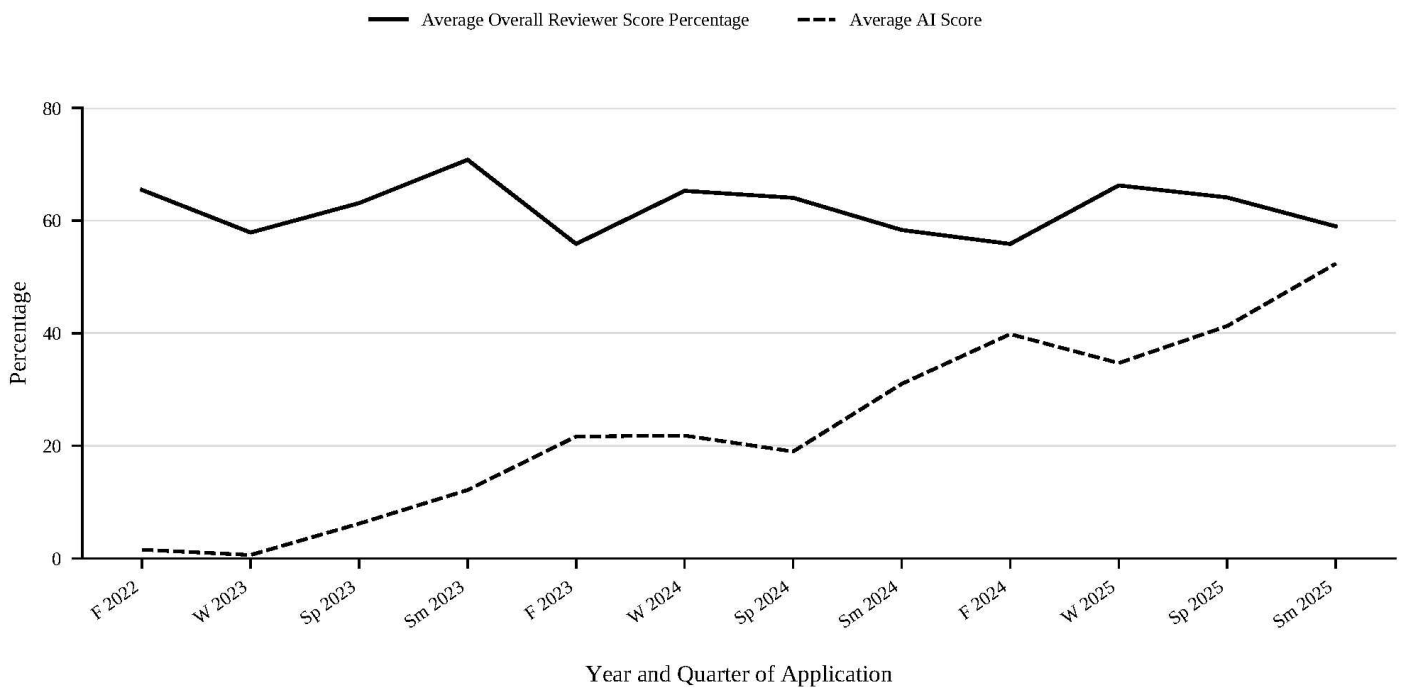


Figure 1

Mean reviewer score and AI score by application quarter and year, 2022–2025. Higher AI scores reflected a higher likelihood that the application was generated using AI

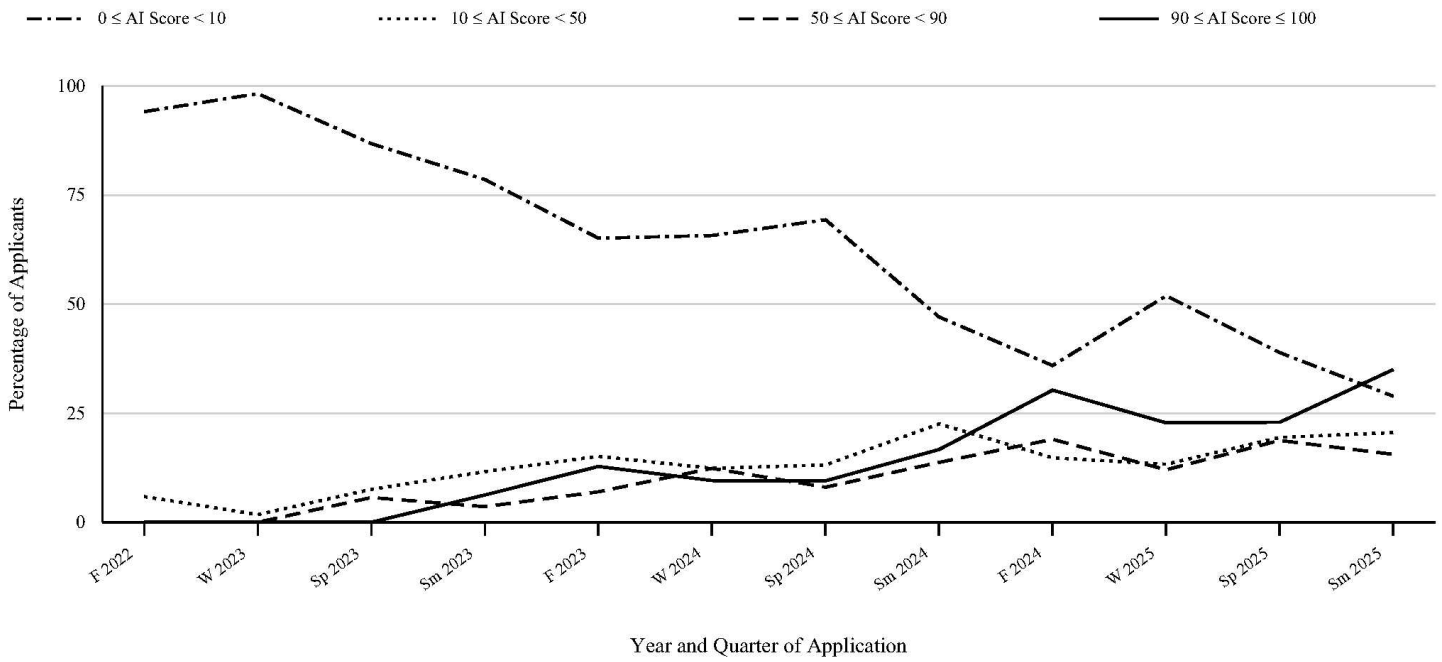


Figure 2

Distribution of AI score categories across application quarter and year, 2022-2025). Higher AI scores reflected a higher likelihood that the application was generated using AI

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionfile1.deidentifiedanalyticdataanddatadictionary.xlsx