(RESEARCH ARTICLE)

Check for updates

Assessing the accuracy of ChatGPT in college admission predictions

Batbolor Urjinbayar [1, *] and Temuujin Tumennasan [2]

[1] Virginia Polytechnic Institute and State University, College of Engineering, Blacksburg, VA, USA.
[2] Independent Researcher, Ulaanbaatar, Mongolia.

Abstract

Generative artificial intelligence (AI) tools are increasingly utilized in higher education, particularly in college admissions. This study assesses OpenAI's ChatGPT in predicting admission outcomes, comparing it to CollegeVine's "chancing" engine. Using 64 Common App submissions from three volunteers applying to various colleges, we asked ChatGPT to predict admission probabilities and decisions based on their profiles. ChatGPT achieved 84.38% prediction accuracy, matching CollegeVine, and had a slightly better Brier Score (0.1164 vs. 0.1186), indicating better probability calibration. These results suggest that generative AI can perform similarly to dedicated admissions prediction models. However, findings are limited to a small sample size, highlighting the need for further research. We discuss the responsible integration of AI in college advising, stressing transparency, fairness, and ethical considerations.

Keywords: Generative AI; College Admission; Decision-Support Tools; ChatGPT; Calibration; Brier Score

1. Introduction

The fast adoption of generative AI in education has introduced both new possibilities and questions for college admissions. The AI tool ChatGPT entered the market in late 2022 and rapidly spread throughout educational institutions and student practices. Research conducted in 2023 showed that about 30% of college students employed ChatGPT for their assignments and 10% of applicants admitted to using the tool when writing their college application essays (Intelligent.com, 2023) [1]. The educational sector started testing AI solutions at the same time colleges implemented AI systems in their admissions offices, reaching more than half of institutions by late 2023 while 82% of institutions planned AI adoption by 2024 (Intelligent.com, 2023) [1]. The adoption of decision support tools (DSTs) in higher education admissions demonstrates the evolution from algorithmic "chancing engines" to AI-driven application screeners.

The accuracy of these tools in predicting admissions outcomes becomes the primary point of inquiry. The decision support tool CollegeVine generates data-based chance predictions through analysis of historical college admissions records (CollegeVine, 2021) [2]. The text-based knowledge allows ChatGPT to make predictions even though it lacks domain-specific information. Our research examines whether ChatGPT demonstrates comparable accuracy to a domain-specific tool in its predictions.

The study evaluates ChatGPT's ability to forecast admission results for 64 applications from three volunteers (Participants A, B, and C). The study evaluates ChatGPT's accuracy through comparison with actual admissions results and CollegeVine's probability estimates by using accuracy percentages and Brier score probability calibration measurements. The research results appear in discussions about AI applications in college admissions (Barnard, 2024; Kiaghadi and Hoseinpour, 2022) [3,4] as well as fairness research (Gayake, 2023; Gándara et al., 2024) [5,6] and ethical

* Corresponding author: Batbolor Urjinbayar

issues (Shermis, 2013) [7]. The research examines the responsible implementation of generative AI systems in college advising practices.

1.1. Positional Statement

Before presenting the methods and findings, we must clarify our researcher and participant positions. Throughout our college application process, we utilized ChatGPT as a navigation tool. We discovered that some classmates used AI to generate essay ideas. This prompted us to explore AI to clarify admission essay prompts and generate a list of scholarships. Our positive experiences with AI tools made us wonder if they could predict admission probabilities with the same accuracy as CollegeVine. This curiosity motivated us to initiate our research project.

It is important to acknowledge that our involvement with AI usage may have introduced several potential biases. Our positive experience could lead me to overestimate the reliability of these tools and overlook their actual limitations. To mitigate personal biases, we employed quantitative methods by providing the same prompt to each model for evaluation and measuring their performance using accuracy and Brier score metrics. Our core concerns include the lack of data transparency in AI predictions and the risk of perpetuating training-based discrimination in data collection.

To address these concerns, our research findings include calibration statistics and an analysis of fairness literature in the discussion section. By outlining our prior experiences, motivations, and doubts, we aim to illustrate how our positionality influences our interpretation of the results

2. Materials and Methods

2.1. Participants and Dataset

This study draws on three volunteer participants, labeled A, B, and C, who collectively applied to 64 colleges via the Common Application. Although each volunteer had a distinctive profile differing in GPA, test scores, extracurriculars, and other factors they all agreed to share their admission results for research. Each volunteer's outcome at each college was recorded as admitted (1) or denied (0) (waitlist decisions were ultimately coded as denies if they did not result in final admission).

2.2. Data per Application

- Volunteer ID (A, B, or C)
- College (private, public, varying selectivity)
- Actual Decision (1 = accepted, 0 = denied/waitlist)
- ChatGPT Probability: ChatGPT's final numeric estimate (0–100%)
- ChatGPT Binary Prediction: Accept if ≥ 50%, Reject if < 50%
- CollegeVine Probability: Probability as output by CollegeVine's DST
- CollegeVine Binary Prediction: Accept if ≥ 50%, Reject if < 50%

2.3. Prediction Tool

2.3.1. ChatGPT

We employed GPT o1. Each volunteer's Common App application and the relevant college name were inserted into a standardized prompt. ChatGPT then returned a numeric chance estimate and a brief textual rationale. We repeated this for all 64 applications across the three volunteers.

2.3.2. College Vine

Each volunteer also input their details into CollegeVine's online calculator, which produces a percentage chance of acceptance for a given college. We recorded this probability for each application in the dataset.

2.4. Evaluation Metrics

2.4.1. Accuracy

$$Accuracy = \frac{Number\ of\ Correct\ Prediction}{Total\ Prediction} * 100\%$$

Predictions were deemed "correct" when the models accept/reject classification matched the actual outcome.

2.4.2. Brier Score

$$Brier\ Score\ = \frac{1}{N}\sum_{i=1}^{N} (p_i - o_i)^2$$

Where p_i is the predicted probability for application i, and o_i is the actual outcome (1 or 0). A lower Brier score indicates better alignment of probability estimates with real outcomes (i.e., Appendix A).

2.5. Procedure

Each volunteer's data was collected via a secure form. ChatGPT predictions were obtained in individual sessions, ensuring no cross-contamination in prompting. The predicted acceptance threshold was set at 50% for both ChatGPT and CollegeVine. We then aggregated all 64 cases, computed overall accuracy (across all volunteers), and calculated Brier scores for ChatGPT and CollegeVine. This allowed us to compare generative AI to a specialized DST on a multi-participant basis, while still preserving the single dataset of 64 total applications.

3. Results

Table 1 Summary of the results

	Chat GPT	CollegeVine
Accuracy	84.375%	84.375%
Brier Score	0.1164	0.1186

Accuracy for both ChatGPT and CollegeVine was 84.38%, meaning neither model clearly outperformed the other in straightforward classification. In terms of calibration, however, ChatGPT demonstrated a marginally lower (better) Brier Score (0.1164) than CollegeVine (0.1186), suggesting its probability estimates were slightly more in line with actual outcomes across these three volunteers' applications. A closer look at individual decisions revealed that both models generally agreed on the likely result. Discrepancies arose primarily in borderline cases, where CollegeVine tended to show more optimism for moderately competitive schools. Occasionally, this optimism proved correct, yet other times it led to overestimations that ChatGPT avoided. Overall, these differences largely canceled out, producing the identical accuracy rates observed.

4. Discussion

4.1. Comparison with Prior Research

Our results align with Barnard's (2024) suggestion that AI could "revolutionize" admissions, as ChatGPT closely approximated a specialized tool's performance [3]. CollegeVine has historically reported robust calibration (CollegeVine, 2021) [2], and in this multi-volunteer dataset, ChatGPT demonstrated equally high accuracy and slightly superior Brier scoring. Kiaghadi and Hoseinpour (2022) noted how a prescriptive analytics approach can enhance admissions decision-making [4]; here, ChatGPT seemingly leveraged textual patterns from its training to emulate such data-driven logic.

4.2. Bias and Fairness

Though the volunteers differ in their personal profiles, three volunteers represent a limited sample, preventing thorough subgroup bias analysis. Gándara et al. (2024) and Gayake (2023) caution that AI may replicate historical inequities, especially if it infers success rates from potentially biased corpora **[6,5]**. Future expansions with more diverse applicants could uncover whether ChatGPT systematically underestimates or overestimates particular groups.

Limitations

- **Small Number of Participants**
  - o Only three volunteers contributed, limiting generalizability.

- **Limited Admissions Cycles**
  - All data reflect only two specific years' outcomes, which may not generalize to future cycles.
- **Model-Specific Behavior**
  - ChatGPT's outputs may vary with prompt design or model updates, reducing reproducibility over time.

Implications and Future Work

For prospective students, these findings highlight that ChatGPT might offer an *alternative or supplemental opinion* to specialized DSTs like CollegeVine. However, reliance on AI predictions for high-stakes decisions remains risky if biases and coverage gaps are not accounted for (Johnson, 2024) [8]. Observing that both AI and DST can converge at ~84% accuracy suggests synergy is possible students may consult multiple sources to gain well-rounded insights into their admissions odds.

Future studies should feature more participants from a variety of backgrounds (Van Busum and Fang, 2024) [9], incorporate fairness audits (Gándara et al., 2024) [6], and analyze how generative AI shapes real user decisions. Explorations could also refine prompts (Shermis, 2013) [7] or feed ChatGPT partial real admissions data to improve calibration.

5. Conclusion

This paper assessed ChatGPT's accuracy in predicting college admissions decisions for three volunteers who collectively applied to 64 institutions, comparing the generative AI's performance against CollegeVine's specialized DST. Both tools matched at 84.38% accuracy, while ChatGPT exhibited marginally better calibration (Brier Score of 0.1164 vs. 0.1186). These results reveal that a versatile AI can approximate a domain-specific model under certain conditions. However, a limited sample size constrains broader claims; admissions processes are multifaceted, subject to policy changes, and often require human context.

As AI's footprint grows in admissions, transparency, fairness, and ethical oversight are critical (Gayake, 2023) [5]. Researchers and practitioners must continue examining potential biases and ensuring that AI-driven guidance complements, rather than supplants, human decision-making. Ultimately, responsibly developed generative AI could enrich the admissions landscape provided we remain vigilant about equity, interpret predictions as probabilistic (not guaranteed), and refine these tools in collaboration with educators and admissions experts.

Compliance with ethical standards

Disclosure of conflict of interest

We have no known conflict of interest to disclose.

Statement of informed consent

Informed consent was obtained from all individual participants included in the study.

References

[1] Intelligent.com. (2023). 8 in 10 Colleges Will Use AI in Admissions by 2024. https://www.intelligent.com/8-in-10-colleges-will-use-ai-in-admissions-by-2024/ [

[2] CollegeVine. (2021). Is CollegeVine's chancing engine actually accurate? [Blog post]. https://www.collegevine.com/blog

[3] Barnard, B. (2024, September 17). College admission: An AI revolution. Forbes. https://www.forbes.com/sites/brennanbarnard/2024/09/17/college-admission-an-ai-revolution/

[4] Kiaghadi, M., and Hoseinpour, P. (2022). University Admission Process: A Prescriptive Analytics Approach. Artificial Intelligence Review, 56(4), 1–24.

[5] Gayake, P. (2023). Bias in AI-Powered Admissions: A Case Study of Historical Data Reinforcement. Artificial Intelligence and Society, 40(1), 89–105. https://doi.org/10.1007/s00146-023-01529-1

[6] Gándara, D., et al. (2024). Inside the Black Box: Detecting and Mitigating Algorithmic Bias Across Racialized Groups in College Student-Success Prediction. AERA Open.

[7] Shermis, M. D. (2013). The Challenges of Automated Writing Evaluation in Higher Education Admissions. In M. D. Shermis and J. Burstein (Eds.), Handbook of automated essay evaluation: Current applications and new directions (pp. 200–221). Routledge.

[8] Johnson, R. (2024). Algorithmic admissions: Ensuring equity in AI-driven decision-making. Brookings Institution Report.

[9] Van Busum, K., and Fang, S. (2024). Bias analysis of AI models for undergraduate admissions in a test-optional era. arXiv preprint.

Appendix A

Table 2 Sample Prediction Comparison for Participant A's Applications

College	ChatGPT Predicted Chance	CollegeVine Predicted Chance	Actual Outcome
Johns Hopkins University	15%	7%	Rejected

Accuracy in this sample Accuracy of ChatGPT

Since the predicted chance is less than 50%, we will consider this as a rejection, which matched the actual outcome. Hence,

$$Accuracy\ (ChatGPT)\ = \frac{1}{1} * 100\% = 100\%$$

Accuracy of College Vine

Since the predicted chance is less than 50%, we will consider this as a rejection, which matched the actual outcome. Hence,

$$Accuracy\ (ChatGPT)\ = \frac{1}{1} * 100\% = 100\%$$

Brier Score

$$Brier\ Score\ (ChatGPT)\ = \frac{1}{1}\sum_{i=1}^{1} \square\ (0.15 - 0)^2 = 0.0225$$

$$Brier\ Score\ (CollegeVine)\ = \frac{1}{1}\sum_{i=1}^{1} \square\ (0.07 - 0)^2 = 0.0049$$