Proceedings of the Twentieth International AAAI Conference on Web and Social Media (ICWSM 2026)

Do Students Still Ask Each Other? Evidence from Reddit on Test Preparation in the Age of Generative AI

Ratanond Koonchanok

Indiana University Indianapolis
rkoonch@iu.edu

Abstract

The growing availability of generative AI tools has drawn attention to their potential influence on how students prepare for standardized tests. We examine whether ChatGPT's launch coincided with changes in SAT-related help-seeking behavior on Reddit using submission posts from SAT-focused subreddits and interrupted time-series analyses. Focusing on requests for peer explanations and strategy-related guidance, we find no evidence of an immediate decline in peer explanation requests overall or by subject, with subject-specific patterns largely following preexisting trends. In contrast, strategy-related help-seeking exhibits a significant increase in post-launch growth, suggesting a shift toward higher-level planning support. Overall, these findings indicate that generative AI did not displace peer-based SAT support but may selectively reshape how students seek help in online communities.

Introduction

Generative AI systems have rapidly entered students' learning workflows, offering instant explanations, worked examples, and personalized guidance for a wide range of academic tasks. Among the many domains influenced by generative AI, education has become one of the most widely discussed (Koonchanok, Pan, and Jang 2024). In the context of standardized test preparation, where repeated practice, step-by-step reasoning, and strategic decision-making are central, tools such as ChatGPT appear particularly well suited to support student learning. At the same time, their growing capabilities raise questions about whether generative AI substitutes for, reshapes, or coexists with long-standing peer-based learning practices.

A growing body of research demonstrates that large language models can perform competitively on standardized and high-stakes assessments. Early studies showed that ChatGPT could achieve strong performance on mathematics components of standardized tests, including a reported success rate of approximately 70 percent on SAT Math-style problems (Dao and Le 2023). Subsequent work found that GPT-4 performs at or near human levels on a range of multiple-choice examinations in higher education (Newton and Xiromeriti 2024), and can generate short-answer and numerical problems that closely resemble university entrance exams (Fernández et al. 2024). Other studies highlight promising performance in reading comprehension and grammatical error correction tasks (Wu et al. 2023; de Winter 2024), as well as broader aptitude and problem-solving assessments (Giannos, Delardas et al. 2023). More recent evidence suggests that ChatGPT-generated instructional support can produce learning gains comparable to those achieved with human-authored tutoring in mathematics (Pardos and Bhandari 2024), and provide effective guidance in domains such as programming (Hartley, Hayak, and Ko 2024).

While this literature establishes the technical capability of generative AI systems to answer questions and support learning, it leaves open an important behavioral question: how does the availability of such tools affect students' reliance on peer-based support? Before the emergence of generative AI, students preparing for standardized tests such as the SAT have turned to online communities to ask questions, request explanations, and exchange preparation strategies. Reddit hosts large, topic-specific forums where students seek help with math problems, reading passages, grammar questions, and broader test-taking strategies. These interactions suggest that peer explanations and strategic advice play a meaningful role in how students learn and prepare outside formal instructional settings.

Whether generative AI alters these help-seeking behaviors remains an open empirical question. On one hand, AI systems capable of providing immediate explanations and tailored guidance may reduce students' need to ask peers for help. On the other hand, peer communities offer contextualized explanations, social interaction, and collective reasoning that AI tools may not fully replicate. Rather than assuming displacement, understanding how generative AI intersects with existing peer-based learning practices requires examining naturally occurring student behavior over time.

In this paper, we examine whether the public release of ChatGPT in late 2022 coincided with changes in how students use online peer communities for SAT preparation. Using Reddit data spanning periods before and after ChatGPT's release, we focus on two common forms of help-seeking. The first is requests for peer explanations of specific SAT questions. The second is requests for strategic advice related to test preparation, such as study planning and

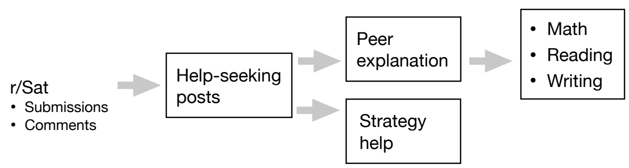

Figure 1: Data processing pipeline

score improvement. We analyze whether the frequency of these requests changes over time and whether patterns differ across SAT subject areas. We address the following research questions:

RQ1: Did the launch of ChatGPT coincide with changes in students' reliance on peer explanations for standardized test questions, both in aggregate and by subject?

RQ2: Did the launch of ChatGPT coincide with changes in the frequency of SAT strategy-related help requests?

By grounding these questions in behavioral data from online communities, this study complements prior work on AI performance by examining how generative AI interacts with peer-based learning in practice and provides insight into whether generative AI represents disruption, reconfiguration, or continuity in how students seek support when preparing for high-stakes standardized exams.

Methods

Data

We collected both submissions and comments from the `r/Sat` subreddit using the Pushshift archive (Baumgartner et al. 2020), with data accessed via Academic Torrents (Cohen and Lo 2014). The subreddit `r/Sat` is an active community where high school students and test-takers discuss SAT preparation strategies, share study resources, and exchange advice on test-taking experiences. We selected August 2021 to July 2024 to align with the academic calendar, avoid COVID-19–related confounds, and ensure a balanced number of observations before and after the intervention.

Identifying SAT Help-Seeking Posts

We begin by extracting SAT-related submission posts that express help-seeking intent using a keyword-based approach, including requests for test-taking strategies and explanations of specific SAT questions, resulting in 42,673 posts. Among the extracted help-seeking posts, we distinguish three mutually exclusive categories: strategy help, peer explanation, and ambiguous cases. Strategy help refers to requests about study plans, pacing, score improvement, or general test-taking approaches. Peer explanation refers to requests for explanations of specific SAT questions or problems. Posts that do not clearly fall into either category are labeled as ambiguous. We first apply a rule-based classification to identify strategy help and peer explanation posts when explicit linguistic cues are present. Because ambiguous cases often require interpretation beyond simple keyword matching, we further classify these posts using a supervised learning approach.

To resolve ambiguous cases, we train a binary text classifier to distinguish between strategy help and peer explanation. We construct a stratified random sample of 200 ambiguous posts for manual annotation, ensuring balanced representation across time. Submissions are categorized as pre- or post-ChatGPT based on whether they were created before or after November 30, 2022, and approximately equal numbers are drawn from each period. These posts are labeled using predefined criteria by the first author and another researcher from the same institution. This stratified sampling approach ensures that the labeled dataset reflects both temporal variation and the diversity of ambiguous posts while remaining manageable for manual annotation. Inter-rater reliability for each assistance type is computed using Cohen's kappa (Cohen 1960) with the resulting kappa value of 0.88. Disagreements were rare and primarily involved ambiguous cases, which were resolved through discussion between the annotators. Using the labeled samples as training data, we train a logistic regression classifier on textual features derived from the submission content. The model outputs a probability score indicating the likelihood that a post requests peer explanation. To prioritize precision and reduce classification noise in downstream analyses, we apply a probability threshold when assigning labels. Posts with predicted probabilities above the threshold are classified as peer explanation, while posts below the threshold are classified as strategy help.

To refine the threshold, we split the training data into a stratified 70/30 train-validation set and evaluate predicted probabilities across multiple candidate thresholds. Predictions within an intermediate range are treated as abstentions to avoid low-confidence assignments. We select a confidence threshold that maximizes precision while maintaining sufficient coverage and apply the classifier to ambiguous submissions identified by rule-based heuristics. Posts falling within the abstention region are discarded. At the end of this step, we obtained 20,897 strategy posts and 13,476 peer explanation posts.

We further classify the peer explanation posts by SAT subject area: mathematics, reading, or writing. We use subject-specific keyword lists to assign posts to each category. Because some peer explanation posts are incomplete or rely on screenshots of questions with minimal accompanying text, certain submissions do not match any subject keywords based on submission text alone. For these cases, we examine the associated comment text and apply the same subject-specific keyword matching. If keywords from multiple subject areas appear within a single post or its comments, we assign the subject based on the category with the highest keyword frequency. This stage resulted in 6,308 math posts, 1,141 reading posts, and 1,628 writing posts. Figure 1 illustrates the overall data extraction pipeline. Details of the post identification process, including the keywords, manual labels, inter-rater reliability score, and classifier, are provided in the supplementary material (https://osf.io/w6kas/).

Interrupted Time Series (ITS)

To estimate the effect of the ChatGPT launch on November 30, 2022, on help-seeking SAT posts, we conducted an

	Strategy		Peer explanation	
	β	p	β	p
Pre-launch trend	-0.009	0.099	-0.017	$^{***} < 0.001$
Immediate level change	-0.028	0.751	-0.075	0.294
Slope change after launch	0.044	$^{***} < 0.001$	0.012	0.102

Notes. β denotes coefficient estimates from interrupted time-series models. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table 1: ITS estimates by help-seeking type

	Math		Reading		Writing	
	β	p	β	p	β	p
Pre-launch trend	-0.016	$^{**}0.001$	-0.021	$^{**}0.001$	-0.016	$^{*}0.038$
Immediate level change	-0.149	$^{*}0.026$	0.045	0.785	0.214	0.185
Slope change after launch	0.017	$^{*}0.021$	0.015	0.317	-0.027	$^{*}0.032$

Notes. β denotes coefficient estimates from interrupted time-series models. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table 2: ITS estimates by subject

interrupted time series analysis (McDowall, McCleary, and Bartos 2019). Because the impact was expected to emerge gradually rather than immediately, we incorporated a ramp function to model the change in trend over time. We include month-of-year fixed effects in all models to account for recurring academic-year seasonality, as posting activity varies across months. We aggregate posts at the monthly level to reduce day-to-day variability. The model was estimated using negative binomial regression to account for overdispersion in the count data. The model specification is as follows:

$$Y_t \sim \text{NegBin}(\mu_t, \theta), \tag{1}$$

$$\log(\mu_t) = \beta_0 + \beta_1 t + \beta_2 \text{post}_t + \beta_3 \text{ramp}_t + \gamma_{\text{moy}(t)} \tag{2}$$

Here, Y_t is the count of posts in month t, μ_t the expected count, and θ the dispersion parameter. time_t indexes the pre-intervention trend, intervention_t equals 0 before and 1 after the ChatGPT release (capturing an immediate level change), and ramp_t is 0 pre-release and increases linearly afterward (capturing gradual post-intervention trend changes). β_0 is the baseline log count, β_1 the pre-intervention slope, β_2 the immediate change, and β_3 the post-intervention slope change. $\gamma_{\text{moy}(t)}$ accounts for recurring seasonal patterns.

A separate control group is not included because ChatGPT's release affected all users simultaneously. Instead, an interrupted time-series design compares post-intervention outcomes to counterfactual trends projected from the pre-intervention period, allowing assessment of deviations from established temporal patterns (Bernal, Cummins, and Gasparrini 2017).

Results

Help-Seeking Type

Table 1 shows distinct temporal patterns for strategy-related requests and peer explanation requests. Prior to the launch of ChatGPT, peer explanation requests exhibit a statistically significant declining trend ($\beta = -0.017$, $p < 0.001$), indicating a gradual decrease in requests for question explanations over time. In contrast, the pre-launch trend for strategy-related requests is negative but only marginally significant ($\beta = -0.009$, $p = 0.099$).

We find no statistically significant immediate level change at the time of ChatGPT's launch for either help-seeking type. The estimated immediate change for strategy-related requests is small and not significant ($\beta = -0.028$, $p = 0.751$), and the corresponding estimate for peer explanation requests is also not statistically significant ($\beta = -0.075$, $p = 0.294$).

Post-launch slope changes differ across the two types. Strategy-related requests show a statistically significant positive change in slope following the launch ($\beta = 0.044$, $p < 0.001$), indicating a reversal of the pre-launch decline and a subsequent increase in strategy-related help-seeking over time. In contrast, the post-launch slope change for peer explanation requests is positive but not statistically significant ($\beta = 0.012$, $p = 0.102$), suggesting no clear departure from the pre-existing trend.

These results indicate that while requests for peer explanations do not show a significant post-launch increase, strategy-related help-seeking exhibits a sustained upward shift following the introduction of ChatGPT. Figure 2 illustrates these patterns by plotting observed monthly counts alongside model-estimated trends. Consistent with the regression estimates, strategy-related requests exhibit a clear upward shift in post-launch slope, whereas peer explanation requests follow a relatively stable trajectory across the intervention, with no visible discontinuity at the time of launch.

Peer Explanation Requests by Subject

We next examine whether patterns of peer explanation requests vary across SAT subjects. Table 2 reports interrupted time-series estimates for mathematics, reading, and writing-related explanation requests.

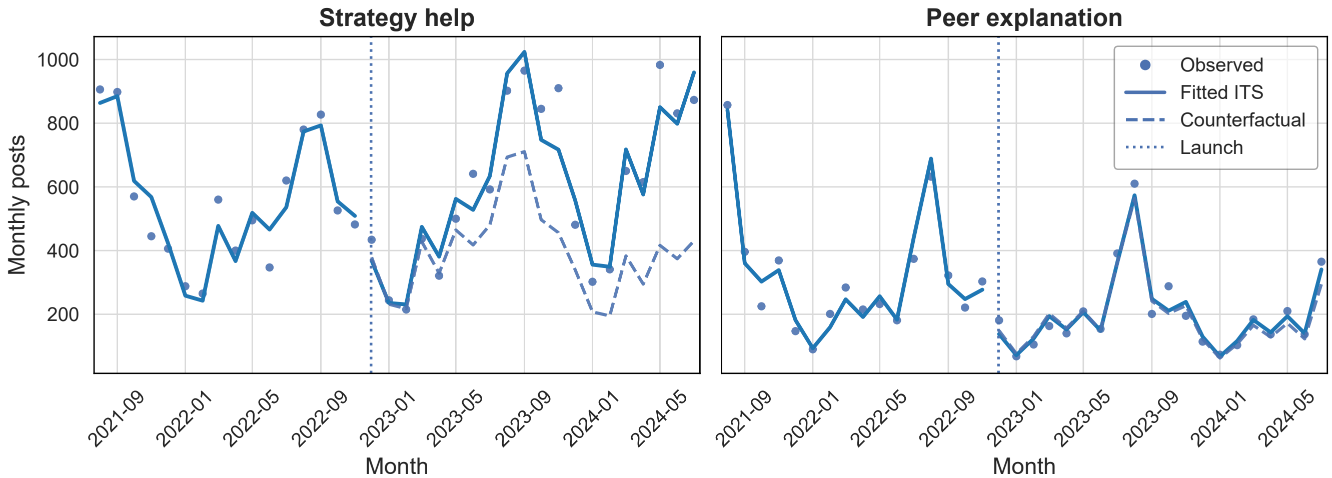

Figure 2: Monthly SAT help-seeking submissions with fitted interrupted time-series trends. Points show observed counts; solid lines indicate model-estimated trends, and dashed lines show counterfactual post-intervention trends assuming no launch. The vertical dotted line marks ChatGPT's public release. Strategy-related and peer-explanation help-seeking are shown in the left and right panels, respectively.

All three subjects exhibit statistically significant negative pre-launch trends, indicating declining peer explanation requests prior to the launch of ChatGPT (Math: $\beta = -0.016$, $p < 0.001$; Reading: $\beta = -0.021$, $p < 0.001$; Writing: $\beta = -0.016$, $p = 0.038$). At the time of launch, we observe a statistically significant immediate decline in mathematics-related explanation requests ($\beta = -0.149$, $p = 0.026$), while immediate changes for reading and writing are not statistically significant.

Post-launch slope changes vary by subject. Mathematics-related explanation requests show a statistically significant positive slope change following the launch ($\beta = 0.017$, $p = 0.021$), indicating a partial reversal of the pre-launch decline. Reading-related explanation requests do not exhibit a statistically significant post-launch slope change ($\beta = 0.015$, $p = 0.317$). Writing-related explanation requests show a statistically significant negative post-launch slope change ($\beta = -0.027$, $p = 0.032$), indicating an accelerated decline after the launch.

These subject-level results suggest heterogeneous responses across SAT domains. While mathematics-related peer explanation requests show signs of recovery following the launch of ChatGPT, writing-related explanation requests continue to decline, and reading-related requests remain largely unchanged relative to prior trends.

Discussion and Conclusion

This study examined how the public release of ChatGPT coincided with changes in SAT-related help-seeking behavior on Reddit, focusing on requests for peer explanations of test questions and strategy-related assistance. Our findings suggest that generative AI did not displace peer-based support in standardized test preparation. While we observe heterogeneous patterns across help-seeking types and subjects, the overall reliance on peer explanations remains substantial, and strategy-related requests show evidence of increased growth following ChatGPT's launch. These results indicate that, rather than replacing peer interaction, generative AI may coexist with and potentially reshape how students engage with online learning communities.

The observed increase in strategy-related help-seeking following ChatGPT's release is consistent with the idea that generative AI can support higher-level planning and reflection while leaving room for human input and discussion. At the same time, the absence of a corresponding increase in peer explanation requests suggests that students may selectively rely on AI for certain forms of assistance while continuing to turn to peers for others. This pattern aligns with prior work highlighting both the potential and limitations of ChatGPT as an educational tool. Although ChatGPT can function as a virtual tutor or instructional assistant, challenges such as generating incorrect or misleading information remain salient, reinforcing the value of peer-based verification and discussion in learning contexts (Lo 2023).

Several limitations should be noted. First, our analysis focuses on what students ask for rather than how requests are linguistically framed; while we distinguish between strategy-related requests and peer explanations, we do not examine differences in wording, depth, or specificity. Generative AI may therefore influence how questions are articulated even if aggregate help-seeking patterns remain stable. Second, we analyze submission posts but not comment-level responses, leaving open the possibility that changes in the quality or style of peer replies occur independently of posting volume. Finally, our findings are limited to Reddit communities focused on SAT preparation and may not generalize to other platforms or standardized tests. Future work could extend this analysis to additional education-focused subreddits (e.g., *r/ACT*) or other testing contexts to assess whether similar patterns emerge across domains and platforms.

Overall, this study contributes empirical evidence that generative AI does not necessarily erode peer-based support in high-stakes educational settings. Instead, online learning communities appear resilient, continuing to play a central role in students' preparation while adapting to the presence of new AI tools. Understanding this coexistence is critical as generative AI systems continue to evolve and become further embedded in educational practice.

Acknowledgments

The author thanks Yanling Pan for assistance with manual labeling and dataset annotation.

References

Baumgartner, J.; Zannettou, S.; Keegan, B.; Squire, M.; and Blackburn, J. 2020. The pushshift reddit dataset. In *Proceedings of the international AAAI conference on web and social media*, volume 14, 830–839.

Bernal, J. L.; Cummins, S.; and Gasparrini, A. 2017. Interrupted time series regression for the evaluation of public health interventions: a tutorial. *International journal of epidemiology*, 46(1): 348–355.

Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1): 37–46.

Cohen, J. P.; and Lo, H. Z. 2014. Academic torrents: A community-maintained distributed repository. In *Proceedings of the 2014 Annual Conference on Extreme Science and Engineering Discovery Environment*, 1–2.

Dao, X.-Q.; and Le, N.-B. 2023. Investigating the effectiveness of chatgpt in mathematical reasoning and problem solving: Evidence from the vietnamese national high school graduation examination. *arXiv preprint arXiv:2306.06331*.

de Winter, J. C. 2024. Can ChatGPT pass high school exams on English language comprehension? *International Journal of Artificial Intelligence in Education*, 34(3): 915–930.

Fernández, A. A.; López-Torres, M.; Fernández, J. J.; and Vázquez-García, D. 2024. ChatGPT as an instructor's assistant for generating and scoring exams. *Journal of Chemical Education*, 101(9): 3780–3788.

Giannos, P.; Delardas, O.; et al. 2023. Performance of ChatGPT on UK standardized admission tests: insights from the BMAT, TMUA, LNAT, and TSA examinations. *JMIR Medical Education*, 9(1): e47737.

Hartley, K.; Hayak, M.; and Ko, U. H. 2024. Artificial intelligence supporting independent student learning: An evaluative case study of ChatGPT and learning to code. *Education Sciences*, 14(2): 120.

Koonchanok, R.; Pan, Y.; and Jang, H. 2024. Public attitudes toward chatgpt on twitter: sentiments, topics, and occupations. *Social Network Analysis and Mining*, 14(1): 106.

Lo, C. K. 2023. What is the impact of ChatGPT on education? A rapid review of the literature. *Education sciences*, 13(4): 410.

McDowall, D.; McCleary, R.; and Bartos, B. J. 2019. *Interrupted time series analysis*. Oxford University Press.

Newton, P.; and Xiromeriti, M. 2024. ChatGPT performance on multiple choice question examinations in higher education. A pragmatic scoping review. *Assessment & Evaluation in Higher Education*, 49(6): 781–798.

Pardos, Z. A.; and Bhandari, S. 2024. ChatGPT-generated help produces learning gains equivalent to human tutor-authored help on mathematics skills. *Plos one*, 19(5): e0304013.

Wu, H.; Wang, W.; Wan, Y.; Jiao, W.; and Lyu, M. 2023. Chatgpt or grammarly? evaluating chatgpt on grammatical error correction benchmark. *arXiv preprint arXiv:2303.13648*.

Paper Checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes
   (e) Did you describe the limitations of your work? Yes
   (f) Did you discuss any potential negative societal impacts of your work? No, we did not see any potential negative societal impacts.
   (g) Did you discuss any potential misuse of your work? No, we did not see any potential for misuse.
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes. Data and documentation are included as part of supplementary materials.
   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes
2. Additionally, if your study involves hypotheses testing...
   (a) Did you clearly state the assumptions underlying all theoretical results? Yes
   (b) Have you provided justifications for all theoretical results? Yes
   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? Yes
   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? Yes

(e) Did you address potential biases or limitations in your theoretical framework? Yes

(f) Have you related your theoretical results to the existing literature in social science? Yes

(g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? Yes

3. Additionally, if you are including theoretical proofs...

(a) Did you state the full set of assumptions of all theoretical results? NA

(b) Did you include complete proofs of all theoretical results? NA

4. Additionally, if you ran machine learning experiments...

(a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? Yes

(b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? Yes, in the supplementary materials, since we used machine learning only as an intermediate step to construct the dataset, rather than as the focus of the paper.

(c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? Yes, in the supplementary materials.

(d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? No because we only use standard libraries and relatively small datasets.

(e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? Yes

(f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? Yes

5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...

(a) If your work uses existing assets, did you cite the creators? Yes

(b) Did you mention the license of the assets? NA

(c) Did you include any new assets in the supplemental material or as a URL? NA

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? NA

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? NA

(f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see **?**)? NA

(g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see **?**)? NA

6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...

(a) Did you include the full text of instructions given to participants and screenshots? NA

(b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA

(d) Did you discuss how data is stored, shared, and deidentified? NA