Perception and Responses of College Students to Financial Scams

Zheyu Huang

May 9, 2025

Research Project Report for Master's of Science degree in Computer Science

Brown University

Introduction

There is an increasing concern regarding youth safety in digital environments, particularly given the rise of conversational AI tools and financial scams targeting youth on social media platforms. A recent BBC report headlined "Character.ai: Young people turning to AI therapist bots"(BBC News, 2024) caught our attention. The piece recounted a 14-year-old's suicide after months of intensive, private chats with a Character-AI bot. This case raises questions on overlooked risks of emotional over-reliance on AI tools that current online safety frameworks largely ignore. Building on that alarm, we pursued a four-phase, mixed-methods investigation to understand these concerns. Initially, we scoped issues relating to youth well-being and emotional dependency on AI chatbots. Subsequently, leveraging our access to the TikTok Research API, we systematically analyzed youth exposure to harassment and scams. The third phase involved cross-platform analyses on Reddit to characterize community responses and scam dynamics further. Finally, we designed and deployed a pilot study as a prelude to our comprehensive survey to gain primary insights into college students' experiences and responses to financial scams. This report documents our findings from each phase, explores implications for youth online safety, and suggests directions for future research and interventions.

Phase 1 - Scoping the issue through Character AI and Youth Well-being

The Character-AI story crystallised two intertwined risks, emotional dependency and unregulated advice, that existing child-online-safety frameworks barely touch. We left that article asking: *how many other teens rely on conversational agents for support, and what unseen harms may follow?*

With that question in mind, we conducted a literature review and apparently two patterns surfaced:

1. **Emotional reliance exceeds clinical intent.** Multiple studies noted that adolescents engage bots for “late-night companionship” far more than for structured therapy, reporting feelings of closeness that blur tool–friend boundaries.

2. **Opaque data practices.** Only one paper detailed what conversational data are retained, and none addressed minors’ rights under COPPA or GDPR-K.

Those gaps came together into an initial guiding question: **“**How does sustained interaction with Character-AI-style chatbots influence young people’s help-seeking behaviour and perceptions of privacy?**”** This framing kept both psychological and informational harms in view.

While gathering real-world chatbot anecdotes on Reddit and Discord, we kept seeing teenagers drop TikTok links about being mocked, doxxed, or “ratio’d” in comments. That organic cross-posting signalled that TikTok was where teens were publicly sharing their harassment experiences**.** Pew’s 2023 teen-media survey backs this visibility: 58 % of U.S. teens visit TikTok every day, with one-in-six there “almost constantly” (Pew Research Center, 2023).

Crucially, our team had just secured university-licensed access to the **TikTok Research API**. That capability made a systematic scrape of harassment content both feasible and timely. Thus we started utilizing TikTok API to scrape user data for more insights.

Phase 2 - Leveraging TikTok Research API

Building on our early investigation it was apparent that TikTok was the dominant platform among youth and with access to the **TikTok Research API,** we developed a modular data pipeline to systematically examine how young people engage with content relevant to harassment. This phase marked a methodological shift from speculative interest in chatbot harms to an empirical investigation of platform-specific risk exposures.

The pipeline evolved across three major iterations. The initial version focused on mastering basic API functionality. We scripted queries that extracted video metadata, hashtags, captions, upload timestamps, and user info using the requests and pandas libraries. This early phase allowed us to test the viability of harassment- and scam-related queries and begin tagging content by thematic relevance.

Surprisingly, our harassment queries unexpectedly surfaced a flood of #TikTokShop refund hacks and “get-rich-quick” schemes among all harassment. This unusual behavior caught our eyes. The large amount of data on harassment related to financial reasons echoed literature on adolescent financial vulnerability. The convergence between high youth engagement and unregulated financial advice prompted a strategic pivot for our research.

With the idea of focusing on financial harassment, in the second iteration of our data processing pipeline, we introduced refined data processing functions, including enhanced parsing logic (parse_content_analysis) and tools for inspecting video samples and behavior patterns

(process_video_data_preview, analyze_results). These updates allowed for more targeted analysis of marketing tactics and user engagement signals—especially those associated with impulsive buying or misleading promotions. We also integrated OpenAI's GPT4o API to analyze both TikTok video descriptions and transcripts. Our prompt instructed the model to generate a concise summary and assess two attributes: whether the content related to financial abuse, and whether it appeared targeted toward teens. This allowed us to detect subtle thematic signals across a large corpus of content that traditional classifiers often missed, such as under-reported harm or adolescent framing. Engagement metrics—likes, shares, and comment volume—were layered onto content metadata to assess virality and impact.

The results were absolutely compelling. Our data suggested that the most common content category among financial harassment is financial scams towards the young population and there are plenty of data points that lead to youth financial vulnerability as well as mental wellbeing issues associated with financial loss.

Phase 3 - Cross-Platform Scam Characterisation

Motivated by the insights found in TikTok, we hoped to find more community discussions and nuanced user-generated narratives to better understand the question. We then decided to adapt our data pipeline to Reddit given its forum nature. We conducted targeted scraping of parent and teenager-focused subreddits. Our data collection specifically included first-hand accounts to preserve authenticity and relevance. These posts were systematically analyzed using a framework that captured persuasion techniques, specific targets of scams, and identified warning signs or red flags of potential financial harm.

Our analysis of Reddit discussions provided preliminary yet crucial insights:

- **Youth-Specific Scams:** Financial scams targeting young users frequently exploited social and emotional vulnerabilities, leveraging peer pressure and aspirational lifestyles as persuasive tools.
- **Parental Intervention:** Parents often sought advice in response to their children's exposure to financial scams, revealing significant gaps in knowledge regarding effective preventive measures.
- **Emotional and Financial Impact:** Emotional distress was frequently cited alongside monetary loss, emphasizing the multifaceted nature of financial scams and their deeper psychological consequences.
- **Community Response:** Peer advice shared within Reddit communities indicated a mixed effectiveness, suggesting a need for structured, evidence-based preventive resources.

Although Reddit data provided rich qualitative narratives and valuable context, it became clear that this approach alone could not quantify the broader prevalence or systematically capture individual experiences and perceptions. This limitation underscored the necessity of acquiring

ground-truth data directly from students to generalize our observations reliably. Therefore, we decided to initiate a comprehensive survey study aimed at college students aged 18 to 24.

Phase 4 - Primary Data Collection via Survey

Having mapped scam tactics **(Phase 2)** and community narratives **(Phase 3)**, we still lacked understanding on how widespread these harms are and why students so rarely speak up. Thus, transitioning into Phase 4, we wanted to investigate private dimensions of how students experience, perceive, and respond to financial scams. To achieve this goal, we developed and refined a detailed survey instrument, designed not only to capture frequency and type of scam encounters but also to delve deeply into personal experiences, emotional reactions, and decision-making processes related to reporting and seeking support.

Building on the insights gleaned from our extensive literature review, we recognized that students aged 18–24 frequently avoid disclosing scam encounters due to embarrassment, fear of judgment, or perceived stigma. Prior studies indicated that such disclosure barriers substantially hinder effective scam prevention and mitigation (Broadhurst et al., 2019; Umar & Dalimunthe, 2024). Recognizing this, we structured our survey to sensitively investigate these private experiences and the associated emotional and psychological barriers to reporting, thereby directly addressing the reporting stigma identified in previous research.

The survey development process commenced with an extensive literature review, drawing heavily on recent academic work and grey literature highlighting key vulnerabilities, psychological impacts, and disclosure behaviors among college students. Preliminary literature established correlations between frequency of social media use, financial literacy, and susceptibility to scams (Umar & Dalimunthe, 2024). We synthesized these insights to draft an initial set of survey questions focused on scam experiences, recognition and response patterns, emotional impacts, and privacy attitudes.

Following initial question drafting, we conducted two rounds of pilot user interviews with college students from Brown and Barnard College to refine our instrument further. These interviews allowed us to test question clarity and relevance, ensuring sensitivity to participants' emotional experiences and potential discomfort. Feedback from these sessions emphasized the necessity of clear anonymity assurances and explicit acknowledgment of discomfort around sharing sensitive experiences—considerations we integrated into our final survey wording.

We then finalized our survey by organizing it into clear thematic blocks, each designed to capture a distinct dimension of participants’ experiences and perceptions. The table below outlines the core components, associated question examples, and the rationale for each thematic area:

Survey overview (Qualtrics, 30-40 min)

Component	Example items / scales	Rationale
Eligibility & Screening	Age ≥ 18, enrolled college student (Q50)	Ensure target population
Scam Exposure	Types × frequency matrix, single "most impactful" incident (Q57, Q3–Q5)	Quantify prevalence & depth
Response & Disclosure	Actions taken (Q13), *reasons for/not reporting* to friends, family, platforms, authorities (Q18–Q27), including shame & fear items	Test stigma hypothesis
Impact Measures	Financial loss bands, emotional distress, reputational harm (Q10–Q12)	Link harms to disclosure choices
Protective Factors	Tech-comfort scale, privacy attitudes (Q36–Q37)	Control variables
Demographics	Gender, age, employment, marital status	Segment analysis

We outlined a targeted sampling strategy, focusing exclusively on college students aged 18–24, aiming for a sample size between 50–200. Participants will be recruited primarily through online platforms and direct college email outreach. Eligible participants will receive a $10 gift card for participation. We are currently at IRB stage and expect the survey to roll out in following weeks.

Reflection

Upon completing data collection, our insights will not end with survey analysis alone. The rich data gathered offers opportunities for impactful interventions. Potential next steps include

co-designing scam-awareness content specifically tailored for platforms like TikTok, leveraging authentic youth engagement styles. Alternatively, we will pitch a “Scam Safety Week” to campus IT and student-life offices, pairing short workshops with QR-coded tip sheets. Longer-term, we envision sharing our annotated dataset and the intervention templates with consumer-protection agencies and TikTok Trust & Safety to develop comprehensive, system-wide preventive measures.

Throughout this project, agility and iterative design emerged as essential methodological frameworks. My research journey began with an extensive literature review, proceeded to quantitative data scraping through API analysis, then transitioned into cross-platform content exploration, and lastly wrote and submitted an IRB-approved comprehensive survey study. Each methodological shift was driven by empirical evidence rather than an expanding scope. For instance, the unexpected prevalence of financial scam content, initially surfaced during queries about harassment on TikTok, prompted us to pivot toward investigating financial scams explicitly. These iterative refinements maintained the responsiveness and relevance of our methods. Furthermore, our methodological breadth significantly contributed to the project's strength. By integrating diverse quantitative approaches including public data mining via the TikTok API, and Reddit content analysis, we obtained a holistic and nuanced understanding of financial scams affecting youth. This comprehensive strategy enriched the accuracy and depth of our insights.

Yet, our study is not without limitations. First, the TikTok Research API is only a partial window into the platform: it withholds video transcripts and creator-supplied tags, hampering nuanced content coding and topic discovery. Second, survey data, although detailed, introduces potential self-report biases such as recall inaccuracies or social desirability bias. Third, the study’s design is strictly cross-sectional. We cannot follow individual participants or creators over time, so behavior change and scam-exposure trajectories remain unknown. Last, our survey sample will remain modest (target $n \approx 200$) and limited primarily to U.S. college students aged 18–24, potentially excluding broader, diverse experiences. Recognizing these limitations illuminates pathways for future investigation, prompting us to consider broader sampling frameworks, longitudinal studies, and expanded intervention strategies that could significantly enhance our impact and understanding of youth vulnerability to financial scams.

Conclusion

Our investigation, iterating from chatbot safety worries to a cross-platform audit of financial scams and, ultimately, to a student-centred survey, revealed deeper threats from financial scams affecting youth across digital platforms. We integrated diverse methodologies to establish a robust foundation to understand the nature and prevalence of financial scams and their psychological implications among college students. We not only mapped the scale of those implications but also surface the social stigma that keeps many losses invisible. We anticipate that our continued research will protect young people’s financial and mental well-being and foster a safer digital environment in an increasingly mediated world.

Reference

1. BBC News. (2024, January 9). Character.ai: Young people turning to AI therapist bots. BBC News. Retrieved from https://www.bbc.com/news/technology-67872693

2. Pew Research Center. (2023, December 11). Teens, social media and technology 2023. Pew Research Center. Retrieved from https://www.pewresearch.org/internet/2023/12/11/teens-social-media-and-technology-2023/

3. Broadhurst, R., Skinner, K., Sifniotis, N., Matamoros-Macias, B., & Ipsen, Y. (2019). Phishing and cybercrime risks in a university student community. *International Journal of Cybersecurity Intelligence & Cybercrime*, 2(1), 4-23.

4. Umar, S. O., & Dalimunthe, Z. (2024). Financial literacy and digital literacy to awareness of investment scams among Indonesian college students. *Eduvest – Journal of Universal Studies*, 4(8), 7215-7227.

5. Tan, K. K. W., Sapiri, H., Md Yusof, Z., & Misiran, M. (2024). Perception of University M undergraduate students in off-campus residential area towards online scamming: A case study. *Perspektif Jurnal Sains Sosial dan Kemanusiaan*, 16(1), 114-128.

6. Ignes, R. N. (2023). Prevalence of online buying scams and fraud exposure among Business Administration students. *The Exponent*, 19(1), 45-58.

7. Trifonova, P., & Venkatagiri, S. (2024). Misinformation, fraud, and stereotyping: Towards a typology of harm caused by deepfakes. In *CSCW EA ’24 Companion* (pp. 533-538). ACM.

8. Tanni, T. I., Akter, M., Anderson, J., Amon, M. J., & Wisniewski, P. J. (2024). Examining the unique online risk experiences and mental-health outcomes of LGBTQ+ versus heterosexual youth. In *CHI ’24 Proceedings*. ACM.

9. Bouma-Sims, E., Hassan, H., Nisenoff, A., Cranor, L. F., & Christin, N. (2024). “It was honestly just gambling”: Investigating the experiences of teenage cryptocurrency users on Reddit. In *SOUPS ’24: Twentieth Symposium on Usable Privacy and Security* (pp. 333-352). USENIX.