Poster: *"Learning Too Much About Me"*: A User Study On the Security and Privacy of Generative AI Chatbots

Pradyumna Shome
Georgia Institute of Technology

Miuyin Marie Yong Wong
Georgia Institute of Technology

1 Introduction

Generative AI has burgeoned in the past few years, leading to highly interactive and human-like chatbots. Trained on billions of parameters and a vast corpus spanning gigabytes of the public Internet, tools like ChatGPT, Copilot, and Bard (which we refer to as generative AI chatbots) write code, draft emails, provide mental health counseling, teach us about the world, and act as mentors to help people advance their careers.

On the other hand, many communities have expressed reservations about widespread usage of such technology. Artists and writers are concerned about loss of their intellectual property rights and the potential for their work to be plagiarized [1, 4]. Educators are concerned about the potential for students to cheat on assignments, and for the chatbots to provide incorrect information [2, 5]. Medical professionals are concerned about the potential for chatbots to misdiagnose patients, and for patients to rely on inappropriate advice [8]. There is fear of the unknown, justified concerns about the potential for misuse, and worry about societal harm. As with other revolutionary advancements in society, there is pressure to adopt these tools to keep up with technology and remain competitive. Before we can bridge this gap, we must understand the *status quo*.

Students are likely to be early adopters of new technology. By examining their initial experiences, we can gain insights into concerns faced by young adults about to enter an AI-integrated workplace. We perform an online survey of 86 students, faculty, and staff at our university, focused on security and privacy concerns affecting chatbot use. We found that participants are well aware of the risks of data harvesting and inaccurate responses, and remain cautious in their use of AI in sensitive contexts, which we unpack in later sections.

Research Question

What security and privacy concerns do students at a large public US university have with adopting generative AI, and how can we overcome them?

2 Methodology

In April 2024, we conducted an IRB-approved online survey with $N = 86$ students, faculty, and staff at Georgia Institute of Technology. The survey was distributed via email to mailing lists of people in various departments. The survey consisted of 28 questions and took approximately 15 minutes to complete. The questions covered a wide range of topics, including the tasks that students use chatbots for, their attitudes towards sensitive queries (medical inquiries, personal and emotional issues, and academics), and their security and privacy concerns. The survey questionnaire is included in the appendices.

3 Results

3.1 Demographics

The median participant was a female-identifying graduate student between 25-34 years of age, with a self-reported proficiency level with technology of 4 out of 5 ("Advanced"), who was moderately concerned about digital privacy.

Graduate students comprised the majority of our survey's respondents. We had 27 participants majoring in Computing, 4 in Design, 23 in Engineering, 10 in the Sciences, 4 in Liberal Arts, and 11 in Business. Participants indicated their level of privacy concern in general and with Generative AI chatbots on a scale of 1 to 5, with 1 being Very Unconcerned, and 5 being Very Concerned. Amongst graduate students, the mean privacy concern was 3.67, and generative AI privacy concern was 3.56. The corresponding figures were 3.40 and 3.08 for undergraduates respectively.

Computing students seem to be the most privacy conscious, followed by students majoring in Design, Liberal Arts, or Business. Engineering and Sciences students are the least privacy conscious. It seems like expert knowledge leads to increased concerns, while some technical knowledge leads to decreased concerns, and presumably lack of detailed knowledge on AI product deployment leads to increased concerns, although not as much as those with expert knowledge.

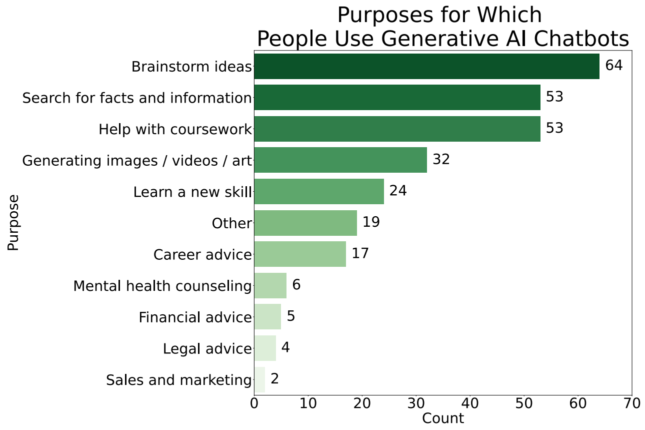


Figure 1: Participant use cases of Generative AI chatbots

3.2 Concerns

Several themes emerged from concerns expressed by students: general concerns regarding data processing practices and accuracy, and domain-specific ones related to medical advice, personal and emotional issues, and plagiarism.

Data Processing Many survey participants are worried about their queries being used to build a profile that could be linked back to them, and the potential for this information to be used against them in the future. For example, people stated that they did not want insurance rates to be affected by their search history, or for their data to be sold to advertisers.

Accuracy Chatbots are prone to responding with incorrect information and generally do not furnish sources of information. When they do not know an answer, they tend to produce an answer that matches the format to the question instead of admit their lack of knowledge, because they are akin to sophisticated text predictors, and incapable of general cognition [3]. Students considered accuracy of responses not only to decide which AI product to use, but also to justify why they did not use a product for some specific types of queries including medical advice and academic assistance.

Medical Advice 50% of participants reported that they would not rely on a chatbot for routine health-related inquiries, whereas 36% expressed ambiguity. People were concerned about compliance with the US Insurance Portability and Accountability Act (HIPAA) [6], and the potential for data collection, storage, and misuse. On the other hand, others mentioned the chatbot's lack of information about existing medical history, potential for misdiagnosis, and possibility of being trained on their inputs, leading to loss of control over private data. There was a strong theme of lack of trust of companies behind the chatbots.

Participants prefer talking to a registered medical practitioner for problems involving reproductive health, chronic conditions, mental health concerns, concerns for which tests or scans need to be performed, and for anything requiring a diagnosis. Users generally restricted their use in this domain to that of a reference that could help them interpret medical jargon, provide general information about a condition, or explain advice given by a doctor.

Personal and Emotional Issues 64% of participants reported their unwillingness with discussing personal and emotional issues with a chatbot. Many people circumvent privacy leakage by asking general questions and treating it as a reference, whereas others ask detailed questions about personal situations without revealing identifying information.

Plagiarism 38% of participants reported using chatbots for help with coursework out of whom 27% were concerned about plagiarism. Most reported using chatbots to help them when they get stuck on questions, produce concept explanations, and least frequently, verify their own solution.

Fact-checking behavior seems to be highly variable, with 3-9 people in every bucket from "Always", "75-100% of the time", "50-75% of the time", and "Less than 25% of the time" to "Never". Students tended to fact-check when querying about a topic they were not familiar with, when the response seemed redundant, inauthentic, conflicting with pre-existing knowledge, or when there were clear errors, such as in mathematics or programming-related tasks.

4 Design Recommendations

We propose several techniques that users and developers of chatbots could implement to mitigate security and privacy concerns.

Privacy Labels Users could be provided with privacy labels [7], a standard set of icons and data fields that clearly and unambiguously detail a given chatbot's data management policy.

Privacy-Enhancing Technologies Chatbots could be equipped with privacy-enhancing technologies such as differential privacy, secure multi-party computation, and trusted execution environments, to prevent leakage of user data.

Prompt Engineering Education Users could be educated on creating prompts that do not reveal personal information, that improve accuracy of responses, and that increase trust in responses. This could be done by providing users with a set of best practices for creating prompts, and by providing them with feedback on the quality of their prompts.

Customizable Language Models Users could be provided with interface affordances that allow them to create chatbots for specific workflows, and have fine-grained control over the datasets they rely on. For example, a city government could create a chatbot that answers questions about city services, and have it trained on a dataset of frequently asked questions, and a dataset of city ordinances. This would allow the city to provide accurate, up-to-date information to its residents, without having to worry about the chatbot leaking sensitive information.

References

[1] Alex Abad-Santos. The writers guild of america is on strike. here's what you need to know. https://www.vox.com/culture/23696617/writers-strike-wga-2023-explained-residuals-streaming-ai, July 2023.

[2] Noor Akbari. The ai cheating crisis: Education needs its anti-doping movement. https://www.edweek.org/technology/opinion-the-ai-cheating-crisis-education-needs-its-anti-doping-movement/2024/02, February 2024.

[3] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the dangers of stochastic parrots: Can language models be too big? 🦜. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, page 610–623, New York, NY, USA, 2021. Association for Computing Machinery.

[4] Andrew Dalton. Hollywood's actors strike is nearing its 100th day. why hasn't a deal been reached and what's next? https://web.archive.org/web/20231020002355/https://abcnews.go.com/Business/wireStory/hollywoods-actors-strike-nearing-100th-day-deal-reached-104136181, October 2023.

[5] Preston Fore. A majority of educators are concerned about how ai may boost cheating and plagiarism but think it will also boost accessibility, according to a new report. https://fortune.com/education/articles/imagine-learning-digital-curriculum-2023-educator-ai-report/, October 2023.

[6] HHS. Hipaa. https://www.hhs.gov/hipaa/index.html, August 1996.

[7] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W Reeder. A "nutrition label" for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security*, pages 1–12, 2009.

[8] Rebecca Pifer. 'hurtling into the future': The potential and thorny ethics of generative ai in healthcare. https://www.healthcaredive.com/news/generative-AI-healthcare-gpt-potential/648104/, April 2023.

5 Appendices

5.1 Survey Questionnaire

Questions marked with an asterisk (*) were required. The survey text has been edited to remove personally identifying information.

5.1.1 Background

I am conducting a study about the experiences of the Georgia Tech community (students, faculty, and staff) with Generative AI chatbots. The survey is aimed at gathering insights into how members of our community interact with and perceive generative AI chatbots, with a particular emphasis on aspects related to security and privacy. The survey is designed to capture diverse perspectives from students (undergrad through PhD), faculty, and staff across various departments. While I cannot instruct participants on what to write, I encourage individuals to share their genuine experiences, thoughts, and concerns regarding their interactions with AI chatbots. Questions center around the varying use cases people consider chatbots for, their usage in the classroom for students and faculty alike, in the medical domain and for personal and emotional issues, and what security and privacy concerns guide their adoption of the technology. This should take about 15 minutes, and not more than 30 minutes.

5.1.2 Study Title

Generative AI Chatbot Experiences

5.1.3 Sweepstakes

- Completing this survey will automatically enter you in the sweepstakes. After the survey and sweepstakes are closed, I will select 10 people at random and send them $10 Amazon.com Gift Cards.

- Note that you don't need to complete the survey to be entered into the sweepstakes. If you would like to enter the sweepstakes without filling out the survey, please submit the form located at Sweepstakes Link.

- Email addresses collected will be deleted after I send the gift cards and will not be used for any other purposes in the study.

5.1.4 Eligibility

At the time of completing this survey, participants meet all of the following requirements:

- Be 18 years of age or older

- Be located in the United States

5.1.5 Consent

By completing the online survey, you indicate your consent to be in the study. You are free to navigate away from this page at any point should you wish to no longer participate, and your data will not be saved.

5.1.6 Benefits

There is no immediate benefit to subjects besides the sweepstakes. It will improve public awareness of privacy concerns a diverse set of people at a large public university in the US have with Generative AI, which could influence their development.

5.1.7 Risks

The risks involved are no greater than those involved in daily activities. You will not benefit for joining this study (unless you are selected in the sweepstakes). We will comply with any applicable laws and regulations regarding confidentiality. To make sure that this research is being carried out in the proper way, the Georgia Institute of Technology IRB may review study records. The Office of Human Research Protections may also look at study records. If you have any questions about the study, you may contact the Principal Investigator by email. If you have any questions about your rights as a research subject, you may contact Georgia Institute of Technology Office of Research Integrity Assurance. Thank you for participating in this study.

5.1.8 Demographics

1. How old are you? *

- Under 18
- 18-24
- 25-34
- 35-44
- 45-54
- 55-64
- 65 and older
- Prefer not to say

2. How would you describe your gender identity? *

- Woman
- Man
- Non-binary
- Prefer not to say

3. How would you describe your proficiency level with technology? *

- Beginner (e.g. experience of less than 6 months with computers or smartphones, regularly need help from more experienced others to complete tasks)
- Novice (6 months or more of experience with smartphones and computers, might frequently ask others for technology related advice)
- Intermediate (a degree of independence using computers without need of regular assistance, may need help troubleshooting problems or performing complex tasks; using computers/smartphones may be small but limited part of your life)
- Advanced (high level of independence using smartphones and computers, can troubleshoot most problems; using computers/smartphones is/was a big part of your life)
- Expert (have degree in computer science or related field, employed in the technology related role such as development/services/administration; comfortable with multiple operating systems; can use command line interfaces)

4. On a scale of 1-5, how concerned are you about your privacy? * *(1: Very unconcerned, 5: Very concerned)*

5.1.9 Basics

Chatbots are computer programs designed to simulate conversation with human users. These have existed for many years in roles such as customer support, but usually had limited options you could select to move the conversation forward. With the advent of ChatGPT, the first chatbot based on Generative AI, in November 2022, chatbots are much more flexible in the manners in which they accept input and how they respond. For example, they can comprehend natural language input in multiple languages, work around typos, and can produce text in various formats, source code for programs, images, and videos.

5. Have you ever used a Generative AI chatbot? *

- Yes
- No
- Yes, ChatGPT 3.5 (OpenAI)
- Yes, ChatGPT Plus (OpenAI)
- Yes, DALL-E (OpenAI)
- Yes, Claude (Anthropic)
- Yes, Bard / Gemini (Google)
- Yes, Stable Diffusion (Stability AI)
- Yes, Poe (Quora)
- Yes, Bing / Copilot (Microsoft)
- Yes, Perplexity AI (Perplexity Labs)

- Other

6. If you haven't used such a chatbot, why not? *(select all that apply)*

- I was not aware of them before this survey.
- I did not want to pay.
- I had security and privacy concerns.
- I didn't know how to use them.
- I didn't think it was relevant.
- Other

7. What makes you choose one chatbot product over another, for a given conversation? *(Be brief, but specific) (free response)*

8. Select all purposes for which you have used a Generative AI chatbot. *

- Searching for facts and information
- Brainstorm ideas
- Help with coursework
- Learn a new skill
- Mental health counseling
- Legal advice
- Career advice
- Generating images / videos / art
- Sales and marketing
- Financial advice
- Other

9. What is your role at Georgia Tech? * *(Select one)*

- Student
- Faculty
- Staff
- Other

5.1.10 Students

Only shown to people who select Student in the previous question.

10. Which of the following best describes your current enrollment status? *

- Undergraduate
- Graduate or Professional
- Non-degree seeking
- Other

11. What college represents your primary affiliation? *

- College of Computing
- College of Design
- College of Engineering
- College of Sciences
- Ivan Allen College of Liberal Arts
- Scheller College of Business

12. Have you used Generative AI in work you have submitted for a class? *

- Yes
- No

13. If yes, how concerned are you of your work being considered plagiarized? *

- Very concerned
- Somewhat concerned
- Neither concerned nor unconcerned
- Somewhat unconcerned
- Very unconcerned

14. How would you describe the level of abstraction used when using Generative AI chatbots in your assignments? *

- I prompt it with the entire description of the task (e.g. the topic of an essay, the description of a program, the problem statement directly).
- I use it to help me when I get stuck on a part of a problem. E.g. a convincing example for an essay, a helper function or new test case in a program, or subproblem encountered while responding to a question.
- I use it to produce concept explanations.

- I use it to verify my work after I have produced my entire response.
- Other

15. How often do you fact-check responses provided via external reliable sources? *

- All the time
- 75-100% of the time
- 50-75% of the time
- 25-50% of the time
- Less than 25% of the time
- Never

16. In what situations do you feel the need to fact-check a response? (free response) *

5.1.11 Faculty

Only shown to people who select Faculty in the previous question.

17. What college best represents your primary affiliation? *

- College of Computing
- College of Design
- College of Engineering
- College of Sciences
- Ivan Allen College of Liberal Arts
- Scheller College of Business

18. As a faculty member, what are your practices relevant to Generative AI in your teaching? *(This could include information about its applications in improving teaching and creating course content, attitudes on student usage of Generative AI in assignments, and more) (free response)*

5.1.12 Social

19. Would you feel comfortable discussing personal or emotional issues with a chatbot? *

- Yes
- No

20. What boundaries would you set when seeking personal advice from a chatbot? *(Personal advice includes but is not limited to topics such as interpersonal conflicts, mental health counseling, and personal development) (free response)*

5.1.13 Medical

21. Would you be comfortable relying on a chatbot for routine health-related queries? *

- Yes
- No
- Maybe

22. What types of concerns would you prefer discussing with a registered medical practitioner instead? (free response) *

23. When you do ask for medical advice, which of the following types of data do you provide? *

- Lifestyle (diet, physical / mental / emotional health practices, etc.)
- Symptoms
- Prior medical history
- Other

24. What privacy concerns do you have when discussing health matters with a chatbot? (free response) *

5.1.14 General

25. Overall, how concerned are you about the privacy implications of using Generative AI chatbots? * *(1: Not at all concerned, 5: Strongly concerned)*

- 1
- 2
- 3
- 4
- 5

26. What privacy concerns do you have about Generative AI chatbots? (free response) *

27. Comment on any pressure you feel to use Generative AI in your personal or professional life. (free response) *

"Learning Too Much About Me": A User Study on the Security and Privacy of AI Chatbots

Pradyumna Shome Miuyin Marie Yong Wong
Georgia Institute of Technology

Motivation

Generative AI Chatbots like ChatGPT, Copilot, and Gemini have disrupted our means of working. Why isn't everyone comfortable with them?

Research Questions

- What are the barriers to widespread adoption and acceptance of Generative AI Chatbots?
- How can users and developers address security and privacy concerns?

Methodology

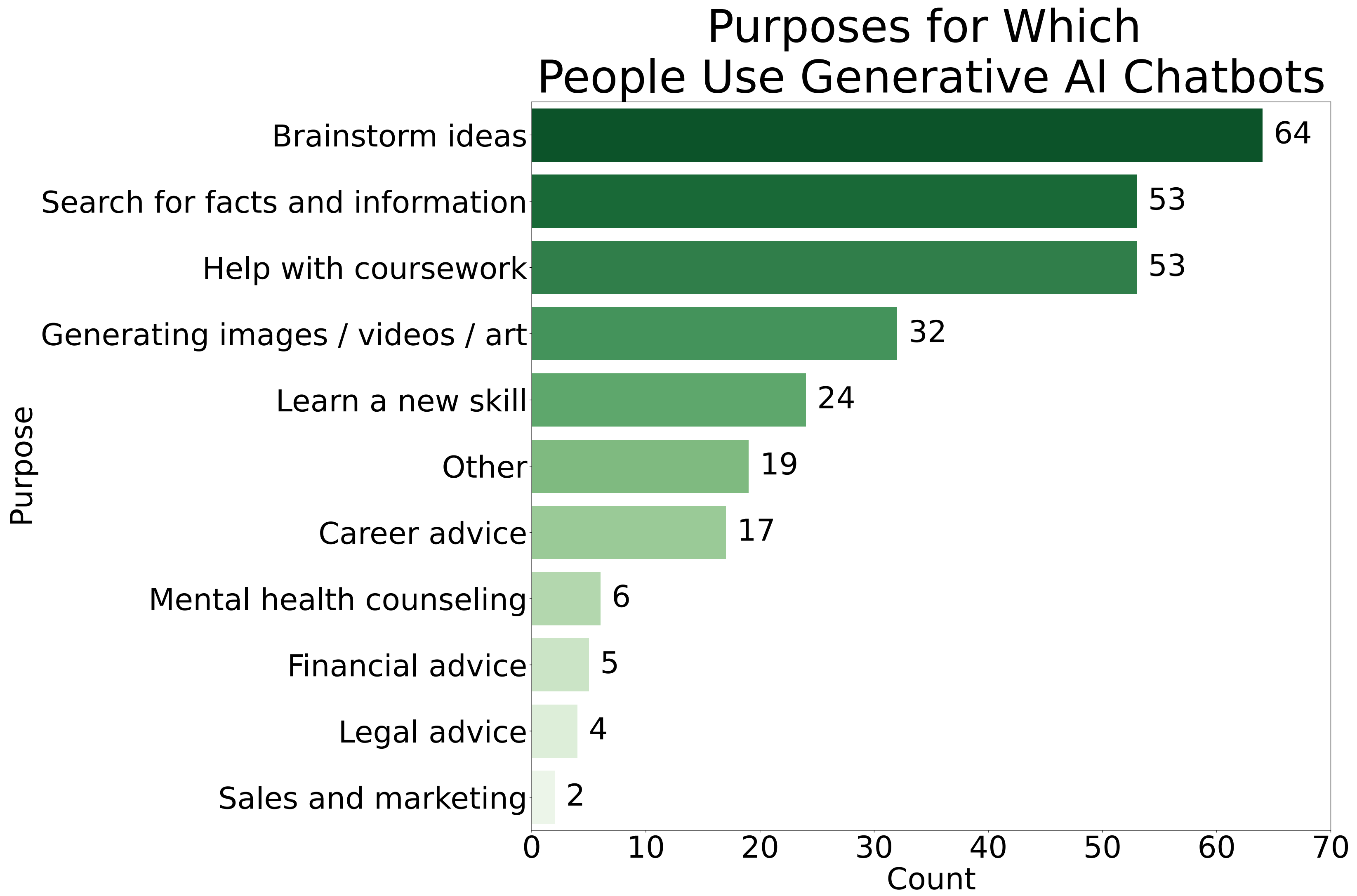
Online survey with *N*=*86* students, faculty, and staff at the Georgia Institute of Technology, focused on experiences with Generative AI chatbots.

Topics addressed: Technology proficiency, levels of digital privacy concern and that of generative AI, products used, use cases, handling of sensitive queries, and use in academia.

Why look at students?

- Early adopters of technology
- Diverse backgrounds and academic / professional interests
- Soon to enter AI-integrated workplaces

Common Tasks



Concerns

- **Accuracy**:
  - 90% fact check responses for academic work.
  - Big reason for non-use in health.
- **Data Processing:** Profile linkability, query privacy, data harvesting, and data breaches.
- **Medical Issues:**
  - 49% of participants will *not* discuss healthcare.
  - Mainly used as a reference, and *not* for diagnosis or complex issues.
- **Interpersonal Situations:** 64% of participants will *not* discuss personal and emotional issues.
- **Coursework:** 17% of students using it for course assignments are concerned about plagiarism.

College	Privacy Concerns	
	Overall	Chatbots
Computing	3.85	3.78
Design + Liberal Arts + Business	3.60	3.40
Engineering + Sciences	3.33	3.09

Table 1. Privacy Concerns by College (1 = Very unconcerned. 5 = Very concerned.)

Notable Trends

- **ChatGPT Plus** Paying users have more use cases.
- **Graduate / Professional Students** Fact-check more often than undergrads and get more writing help.
- **Faculty Usage Scenarios** Generating ideas for curriculum and assignments, reviewing literature, and increasing student exposure to AI.

Design Recommendations

- **Privacy Labels** Standard labels for data use policy
- **Privacy Enhancing Technologies** Differential privacy, secure hardware, and secure MPC
- **Prompt Engineering Education** Redacting secrets, asking for steps to validate results provided
- **Customizable Language Models** High integrity training data for regulated domains such as healthcare, finance, and law