Unveiling Public Sentiments Surrounding Large Language Models (LLMs)

Folakemi Shofu, *Washington University in St. Louis* Ben Ko, *Washington University in St. Louis*

Abstract

The emergence of Large Language Models (LLMs) has drastically transformed our lives in various domains including but not limited to education, healthcare, software engineering, video production, etc. However, questions remain on general guidelines for LLM development as the developing parties are often detached from the general audience. In this paper, we provide a comprehensive overview on public perception surrounding LLMs and security risks that may emerge in various domains where LLMs are already utilized or have potential to be utilized, which could then be used as a guideline for LLM developments.

1. Introduction

With the rise of ChatGPT and its publicity, LLMs have been utilized in our everyday domain more than ever. However, it is a common knowledge that the model development is detached from the public as they are just in the hands of a handful of large corporations and their developers.

In this paper, for the model providers to better reflect public opinion on LLMs, we aim to provide public perception surrounding LLMs through three research questions (RQ) that we try to answer through public survey.

- **RQ1. What do the public think of LLMs?** We try to gauge how well users trust outputs of the LLMs or LLMs themselves and how the models are utilized to perform tasks in multiple domains.
- **RQ2. In what ways will LLMs be used in the future?** We asked laypeople how the future of LLMs would be and where they would be in the future. As the survey participants' backgrounds are multiple domains, we expected the research question
- **RQ3. What security risks/vulnerabilities will arise and how can they be mitigated?** We tried to gather security risks related to LLMs from a layperson perspective of multiple domains to gather novel perspectives.

2. Related Work

We draw inspirations from previous works that conducted high-level studies on public opinions on LLMs and vulnerabilities of LLMs.

Particip-AI [1]. One of the greatest inspirations of this study was Particip-AI [1], a survey framework designed to gather public opinions on development of AI and encourage democratic AI. Our work aims to provide more in-depth views of public on LLMs, which has been exemplified with the emergence of ChatGPT.

Gender bias in LLMs. Recent work [2] demonstrates how implicit and explicit gender biases are present in widely-used LLMs. By reflecting public opinions on development and training process of LLMs, we expect these issues to be mitigated.

Security and privacy challenges of LLMs. As a relatively new technology, new security and privacy challenges of LLMs are emerging [3] and developing solutions have been another aspect of concern in the field. Our work aims to gather novel perspectives on the challenges from diverse backgrounds.

3. Methods

We now describe the methodology of how we designed and conducted the survey to collect public views on LLMs and relevant questions to answer our research questions. We also describe the data analysis methods we used.

3.1 Data Collection

3.1.1. Survey Method

We used Google Forms to construct the survey and distributed the survey through emails for online distribution and with flyers of QR code redirecting to the survey for offline distribution. We promised the respondents to be entered into a $30 draw to compensate the users for their participation. The emails containing link to the survey were distributed to all students, faculties, and teaching assistants enrolled in Washington University (WashU)'s CSE 527S, which ranges from undergraduate students to PhD students, other WashU students with personal connections (enrolled in same classes, friends, teaching assistants, etc.), Chancellor Andrew D. Martin (whom we did not hear back from), and several WashU Professors. We also distributed flyers with QR code to potential respondents that we encountered on Danforth Campus. The survey was open from March 28th. 2024 and closed on April 11th, with the survey collection duration being 14 days. We received 78 responses in total.

3.1.2. Survey Outline

The survey was crafted to capture both quantitative and qualitative data. Specifically, questions Q8-SQ15 were

selected to prompt users to identify risks and vulnerabilities from different perspectives. The survey was divided into seven sections with SQ presented in the paper being an acronym for Survey Question. Provided below is the outline of the survey:

1. **Informed consent form and goal of the study**: In the beginning of the survey, we disclosed the goal of the study and noted that users were free to exit study at any time.
2. **Overview of LLMs**: We provided brief background information on LLMs and their training processes. We also provided common examples of LLMs including ChatGPT, Sora, and Custom GPTs.
3. **Background (SQ 1-5):** We asked the survey participants to provide their education and experience with LLMs.
4. **Future of LLMs (SQ 6, 7):** Before asking the questions on how they think LLMs would be used in the future, we first provided current use cases of LLMs in common daily lives including ChatGPT, search engines, language translation, and chatbots used in customer services. We also made an additional note for respondents to not use AI generated answers for better reflection of public perception of LLMs.
5. **Risks of LLMs (SQ 8-16):** Before asking the questions on the risks and vulnerabilities of LLMs, we first provided a few examples of common attack techniques against LLMs and the attack scenarios. The presented techniques were prompt injection, data poisoning, and jailbreaking. We also provided a screenshot of an actual prompt example of LLM jailbreaking attack, DAN (Do Anything Now). We proceeded with SQ 8 to 16 after these examples.
6. **Contact Information for $30 draw.**
7. **Contact Information for follow-up interview.**

3.1.3. Survey Questions

- **SQ1**. What is your field of study/major/profession? (Please utilize "Other..." option and enter accordingly if there are no available options below or you are not clear about the options or you believe the options may not accurately reflect your response)
- **SQ2**. If you are pursuing an undergraduate degree, what year are you in? If you are pursuing a graduate degree (including dual degree students) what is the degree, you are pursuing? If you are a faculty, what is your position? (Please utilize "Other..." option and enter accordingly if there are no available options below or you are not clear about the options or you believe the options may not accurately reflect your response)
- **SQ3**. How would you rate your familiarity with LLMs?
    - Respondents were given an option to self-evaluate their familiarity with LLMs in a range of 1 to 10 where 1 was presented as “I didn't know what LLMs were until this survey” and 10 presented as “I can conduct original research on LLMs.”
- **SQ4**. How well do you trust LLMs?
    - Respondents were given an option to self-evaluate their trust level of LLMs in a range of 1 to 10 where 1 was presented as “I have no trust in LLMs” and 10 presented as “I have complete trust in LLMs and would be okay providing sensitive information”.
- **SQ5**. If you have any experience working with/using LLM, please list/describe any of them (Common day-to-day use cases include: smart speaker, utilizing ChatGPT, Google Translate, image generation service, etc.).
- **SQ6**. In what area do you think LLMs will be most used in the future? (e.g., education, healthcare, entertainment, programming, content generation, law etc.)
- **SQ7**. What specific tasks or functions do you envision LLMs being used for in that area in the future (other than those mentioned above)?
- **SQ8**. Given your academic/professional background, what security vulnerabilities or privacy risks do you think might arise from the increased reliance on LLMs in everyday tasks and interactions in your field.
- **SQ9**. In your opinion, what are some possible ways that LLMs could be exploited for malicious purposes in the future, given their ability to generate highly convincing and human-like content?
- **SQ10**. In what ways might LLMs compromise user privacy or security, and what measures can be implemented to protect sensitive information?
- **SQ11**. What other risks do you think could arise from the reliance of LLMs in your field in the future?
- **SQ12**. Do you see any specific industries or sectors that might be particularly vulnerable to attacks leveraging LLMs in the future?"
- **SQ13**. In what ways can interdisciplinary collaboration between experts in AI, ethics, law, psychology, and other fields help identify and mitigate the risks associated with LLMs?
- **SQ14**. Do you believe LLMs should have the capability to fingerprint users and identify them based on their inputs? If yes, do you foresee any

potential privacy or security risks associated with this feature?

- **SQ15**. With the development of AI models that can understand text, images, and audio together, do you think there could be any additional security concerns? List those concerns below.
- **SQ16**. How well do you trust LLMs now?
    - Respondents were given an option to self-evaluate their final trust level of LLMs after completing the survey in a range of 1 to 10 where 1 was presented as "I have no trust in LLMs" and 10 presented as "I still have complete trust in LLMs and would be okay providing sensitive information".

4. Evaluation

We organized our analysis into 3 main categories that align with our 3 research questions and discuss the results and findings in this section.

4.1 RQ1. What does the public think of LLMs? (SQ 1-5, 14, 16)

4.1.1. Familiarity.

We asked participants to rate their familiarity with LLMs on a scale of 1-10 and 67.9 % of the participants rated their level of familiarity to be 5 or below. This suggests that a significant portion of participants had limited exposure or understanding of LLMs prior to the study.

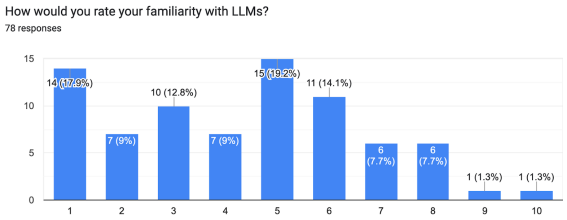

Figure 1. The self-reported familiarity with LLMs returned a median of 5.0 and mean of 4.24 which led us to believe that our survey respondents are a good representation of layperson users of LLMs.

4.1.2. Initial Trust

Like the trend observed in Familiarity, 69.1% of the participants picked a score of 5 and below for their level of trust. However, the distribution of level of trust and familiarity differ significantly (*see figure 1 and figure 2*). Notably the participant with the highest level of familiarity (10) had the least trust (1) in LLMs. However, amongst participants with high familiarity levels, opinions were divided: some indicated low trust (scores <= 5), while others indicated high trust (scores > 5). This discrepancy, coupled with 79.1% of participants with highly familiar reporting interaction with ChatGPT as their main exposure to LLMs in SQ5, suggests that reported familiarity might reflect perceived rather than actual familiarity.

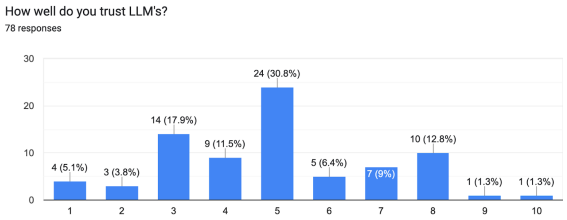

Figure 2. The self-reported trust level with LLMs returned a median of 5.0 and mean of 4.85 which led us to believe that survey participants show neutral trust to LLMs.

4.1.2 Initial Trust by Field of Study

To address the uneven distribution of participants across different fields of study (see figure 3), we analyzed the average level of trust within each field. We discovered that participants studying business showed the highest level of trust in LLMs, with a mean greater than 5, while participants in the field of nursing trust LLMs had the least. This could suggest that individuals in fields where the risks associated with LLMs have immediate and obvious effects (such as healthcare, where incorrect output of LLMs could greatly affect patients and human lives) tend to have less trust in LLMs.

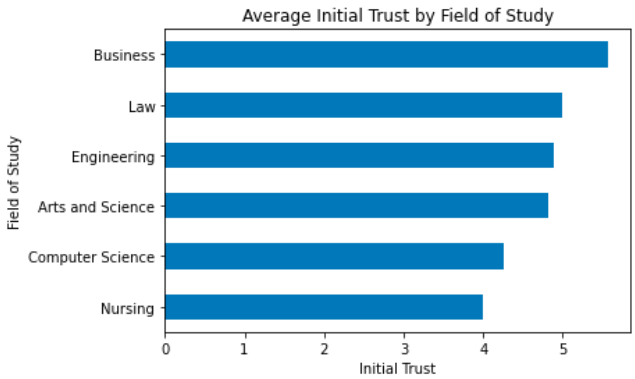

Figure 3. Average initial trust level by field of study.

4.1.3. Final Trust

At the end of the survey, we reassessed trust in LLMs to determine if the overview of LLMs and considerations on their security and privacy implications within the survey would impact their trust. The average trust score decreased from 4.85 to 3.99. In addition, compared to the distribution

of initial trust, the distribution for final trust clustered towards the mean with most participants giving a rating between 3 and 4. (*See figure 4*) This implies that the survey may have negatively impacted the trust level of LLMs as participants themselves think of vulnerabilities and possible exploitations of LLMs.

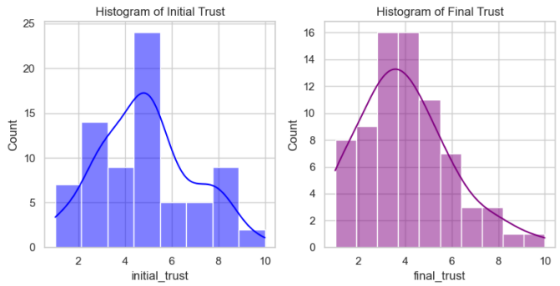

Figure 4. Initial trust level at the start of the survey and final trust level after survey

4.1.4. Sentiment on adopting fingerprinting technique to LLMs

In SQ14, participants were asked whether they feel like LLMs should be able to fingerprint users, an idea currently being discussed in the field. Fingerprinting refers to a technique that could identify users through collected data without explicit form of identification. A sentiment analysis was performed on their responses using HuggingFace sentiment analysis package, on a scale of 1 to 5 where higher number indicates more positive sentiment. Figure 5 suggests that the prevailing sentiment is unfavorable to the idea of fingerprinting.

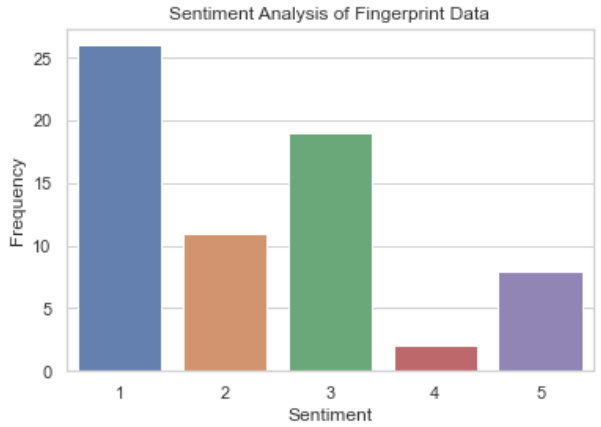

Figure 5. Sentiment distribution of fingerprinting technique in LLMs

4.2 RQ2. In what ways will LLMs be used in the future? (SQ 6, 7)

4.2.1. Areas that are most likely to adopt LLMs

Participants identified several areas likely to adopt LLM usage in the future. These areas were further classified into 8 categories Programming, Education, Healthcare, Entertainment, Research, Law, Business and Other. The Entertainment category received the most mentions (40), followed by Programming (24) and Healthcare (23).

4.2.2. Tasks that are most likely to utilize LLMs

In response to SQ7, tasks that fall under software development and content generation were referenced the most. Other anticipated usages fell under these categories: Automated Tasks, Content Generation, Education Assistance, Healthcare Assistance, Legal Assistance, Social, Programming, Marketing and Advertising, Human Resources, Research. These broad themes referenced by participants likely reflect an interest in harnessing LLMs to drive efficiency and innovation across different sectors.

	Main Category	Frequency	Category
0	Automated Tasks	17	brainstorming, candidate screening, customer service, grammar editing, pattern matching, perform routine tasks, scheduling, summarizations, task summarization
1	Content Generation	23	content creation, content generation, create designs beyond human imagination, creating architectural plans, creating kids bedtime stories, generating animations, generating content, generating lesson plans, generating movies, generating scripts, generating workout plans, language translation, music generation, news article generation, speech writing, writing jokes
2	Education Assistance	4	create lesson plans, homework assistance, study aide
3	Healthcare Assistance	12	analyzing medical scans, diagnose patients, patient diagnosis, patient interaction, patient interaction analysis, phi access, telehealth, virtual therapist
4	Human Resources Tasks	3	interviewing, mitigating bias
5	Legal Assistance	2	legal research
6	Marketing and Advertising	3	personalized advertising, targeted advertising
7	Other	15	all, generating novels, increasing human awareness, mitigate research bias, optimize data, study tool, tasks oraganization, translation, urban planning, virtual reality
8	Research Assistance	9	building financial models and financial research, finding resources, information queries, mapping tumors, research, research assistance, research synthesis, targeted research
9	Social	2	virtual companion, virtual human interactions
10	Software development	27	adaptive programming assistance, advanced code generation, automated writing, creating models, debugging code, generating code, migrating legacy codebase

Figure 6. Tasks that are likely to utilize LLMs categorized

4.3 RQ3. What security risks/vulnerabilities are present/will arise and how can they be mitigated? (SQ 8-15)

4.3.1. Potential exploitations

Participants identified several ways LLMs could be exploited for malicious purposes in the future, given their ability to generate highly convincing and human-like content. Primary concerns included the use of LLMs to perpetrate scams, create deep fakes, spread misinformation, and impersonate people. Particularly troubling despite not being a primary concern for most participants, one participant mentioned the ability for LLMs to be used to exploit human vulnerabilities like loneliness. Moreover, Social Engineering was referenced multiple times outside of scamming, emphasizing participants' concern that scammers may be leveraging LLMs specifically for psychological manipulation. Additionally, participants made multiple references to political disinformation outside the scope of general

misinformation, emphasizing the concern that LLMs may be used to influence the upcoming political elections.

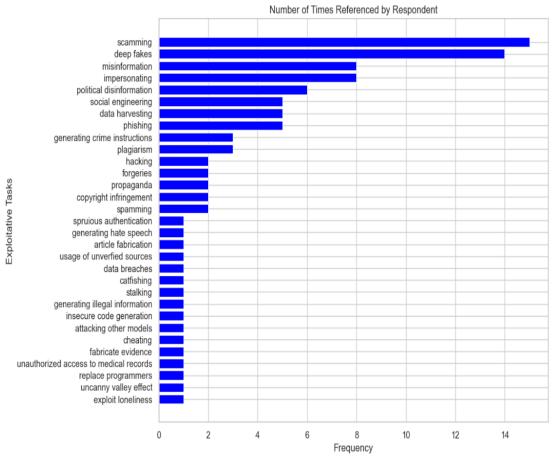

Figure 7. Frequency of exploitative tasks mentioned in survey responses

4.3.2. Most vulnerable areas

Participants identified several sectors likely to be vulnerable to attacks leveraging LLMs. These sectors were further classified into 10 categories: Technology, Arts and Media, Business and Marketing, Crime, Education, Finance, Government and Military, Healthcare, Research and others. The Arts and Media category was referenced the most by participants (18 times) and categories mentioned intersect with explorations related to misinformation, disinformation, deep fakes and copyright infringements forgeries. Additionally, healthcare and the government were also noted as highly vulnerable sectors, with most concerns about LLM usage in those fields related to data privacy and security, particularly the catastrophic effects of sensitive data falling into the wrong hands due to a data breach of LLM data stores or a type of prompt injection.

Main Category	Frequency	Category
Arts and Media	18	arts and entertainment, fashion, journalism, media
Business and Marketing	10	advertisement, customer service, finance, marketing, sales
Crime	2	criminal justice, social engineering
Education	11	education
Finance	2	banking, insurance
Government/Military	11	government, military, politics
Healthcare	13	healthcare
Other	3	manufacturing, social media
Research	2	research
Technology	11	cybersecurity, data science, software development

Figure 8. Sectors leveraging LLMs that may be vulnerable according to survey responses

4.3.3. Risks due to increased reliance, data privacy risks, and security risks

Even when prompted for risks that may arise due to increase reliance, many participants still indicated risks related to privacy and security. This indicates that a primary concern for participants is the mishandling of sensitive data which could lead to data harvesting and brokerage. The inclusion of LLM data into the data brokerage industry would have implications that extend beyond traditional privacy concerns. LLMs that possess the ability to fingerprint users could generate highly detailed user profiles that significantly enhance targeted advertising and could be used to compromise current audio and video identity verification mechanisms like voice and facial recognition.

	Main Category	Frequency	Category	Percent
0	Data security and privacy	41	compliance with data protection laws, data harvesting, deep fakes, encryption, intellectual property ambiguity, intellectual property theft, misinformation, privacy breaches, selling user data, usage of unverified information	46.6%
1	Education and research	9	academic integrity, bias in research data, decrease in human creativity, fabrication of research data, overreliance on ai in education, overreliance on technology	10.2%
2	Healthcare	7	hipaa violations, patient health information leaked, protecting patient files, unsafe medical advice, untested health care	8.0%
3	Legal and compliance issues	3	impersonation and scams, legal research complexities	3.4%
4	Other	16	biased output, compromised data analysis. misinformation, data poisoning. compromised data analysis, data selling, impersonation, incorrect output, jail breaking, misinformation. patient health information leaked, overreliance on technology for routine tasks, phishing, prompt interjection, scam calls, sensitive data analysis performed public servers, spamming, stalking, usage of unverfied sources	18.2%
5	Programming	12	data poisoning, debugging issues, insecure code generation, jailbreaking, prompt injection, security vulnerabilities in coding, skewed training data, unsecure code generation	13.6%

Figure 9. Risks from increased reliance

	Main Category	Category	Solutions
0	Other	corporate espionage, data leekage, data storage, leveragig llm's to generating password combination, unsecure data trnsmission	restrict llm usage for companies handling sensitive information
1	consent issues	data harvesting, data sale, generation of user profile based on chat history, model retraining using user data, selling data to brokers	regulations for optimizing models using user input, legislation, delete sensitive history
2	data breach	data breach, hacking llm data stores, jailbreaking to reveal proprietary data, prompted injection	encryption for data sent between llms, implement robust llm security measures, anonymized sensitive data storage, localize llm usage
3	invasions of privacy	chat history leakage, exploitation of present visual and audio identity verification methods, identity theft, leveraging llms to generate password combinations, subpoena of user history	identity verification before registration, warn users of the risks of inputting sensitive information, restrict llms to non-sensitive tasks, training on pii or being given access to pii for augmentation, training users on llms to ensure safe interactions

Figure 10. Privacy and security risks

4.3.4. Other risks of LLMs

Participants were also prompted to think of other risks besides those related to increased reliance and privacy and security. All risks were aggregated into a single comprehensive list, with each risk categorized into the groups targeted: Individual, Society and Industry. *Figure 11* provides a detailed breakdown of these. For individuals, participants indicated concern about the potential of social and cognitive decline due to overreliance on LLMs for problem-solving and decision-making. Broader societal concerns include the potential for biased content generation, impact to the nature of research itself if journals are flooded with AI written articles, loss of human interaction due to virtual companionship, and the spread of misinformation. From an industry perspective, participants seem concerned about data breaches, loss of jobs due to AI substitution, and decreased strategic decision-making.

Category	Risks
Individual	false or wrong medical diagnosis, identity theft, unconsented data collection, deep fakes, overuse of AI disrupting patient-doctor relationship, model retraining using user data, generation of user profile based on chat history, leveraging llms to generate password combinations, subpoena of user history, biased medical diagnostics harming minority groups, intellectual property theft, phishing, misinformation, spamming, stalking, incorrect model output, exploitation of visual and audio identity verification methods
Society	loss of human interaction, dependency on LLMs leading to social and cognitive decline, loss of human critical reasoning skills, results becoming unmotivated and decontextualized, flooding journals with fake articles, generation of fake or biased content, misinformation, loss of human interaction, biased medical diagnostics harming minority groups, impersonation and scams
Industry	data breaches, corporate espionage, insecure data transmission, selling user data, loss of jobs due to AI substitution, decrease in strategic and creative business decisions, overreliance on AI leading to less competent employees, proprietary data being stolen, automated vulnerability testing used for hacking, data poisoning, prompt injection, data leakage, security vulnerabilities in coding, insecure code generation, debugging issues, jailbreaking, skewed training data, legal research complexities, disservice to the development of young lawyers, job displacement of support staff, data harvesting, selling user data, hipaa violations, patient health information leaked, untested health care, unsafe medical advice, overreliance on technology for routine tasks

Figure 11. Other risks associated with LLMs categorized into risks related to individuals, society, and industry

4.3.5. Risks associated with Multimodal LLMs (MLLMs)

SQ15 gathered users' perception on multimodal LLMs that has capacity to take multimodal inputs like Ferret [4] and/or generate multimodal outputs like Sora. As shown in Figure 12, one of the responses to the question was Deepfakes. While this may indicate that participants may not have distinguished other generative AI technologies from LLM, LLMs and deepfakes are closely related. A recent study [5] suggests integration of LLMs to already existing deep learning tools could enable lay users to create Deepfakes much easily as the creation of fake dialogues can be done in multiple languages using LLMs. [5] also shows current use cases of LLMs being integrated with other deep learning tools to output synthetic videos. Since MLLMs would only get better from here, we believe this may result in another serious issue in the near future.

Figure 12. Word cloud of risks associated with MLLMs

5. Ethics

Our study anonymized every response and collected email addresses only for draw for the compensation of the study. Participants were also notified of the goal of the study prior to survey and were free to withdraw at any moment of the survey. The participation in follow-up interviews was completely voluntary. We do not disclose any personal information of the survey participants in our data as well.

6. Limitations

Ambiguous definition of trust level. For both SQ4 and SQ16, we received feedback from Professor Iqbal that the question may have led the survey participants to interpret it in two different meanings: 1. Whether the respondent trusted in the output of the LLMs to be correct 2. Whether the respondent trusted the LLM-based systems in the context of privacy. To mitigate the issues, we sent out follow-up emails for clarification of how the respondents interpreted the question. However, as we did not gather contact information of the respondents initially, the survey participants we could reach out to were limited. We only heard back from two respondents, which did not help in clarifying how the participants interpreted the question. We plan to design our questions to be more clear in the future.

How far is the future? Another feedback we received from Professor Iqbal was that the survey questions asking about the future use cases resulted in mostly responses that are shown in the present use cases of LLMs. He pointed out that it may be a result of the question being interpreted as the "near" future, where the use case may not differ significantly from now. He suggested the questions be revised to specify how far the future will be, e.g. "In what area do you think LLMs will be most used in 2100?" instead of "In what area do you think LLMs will be most used in the future?". We plan to incorporate the feedback in future studies.

Size and demographics of survey participants. One of the major limitations of our work is that the size of the respondents and the demographics are limited. The survey size of 78 respondents may not be a good reflection of how the general public think of LLMs. Furthermore, most of the survey respondents were undergraduate students (73.3%), which will not be a good representation of various age ranges.

Survey question not capturing participants' backgrounds. We obtained information about the colleges that the survey participants are affiliated with. However, we overlooked the opportunity to gather additional details such as respondents' majors, which could have provided a more comprehensive understanding of their perspectives. Restricting our data collection to only the colleges, e.g. "Arts and Science" instead of listing of majors, at WashU failed to capture the diverse range of respondents who were not affiliated with university and therefore lacked the understanding of the equivalent college, as well as the various majors within those colleges.

7. Discussions and Conclusion

In this study, we conducted a comprehensive survey to investigate public sentiments regarding Language Models (LLMs) and the associated security risks across diverse domains. Our findings reveal a nuanced landscape, with respondents expressing a generally neutral level of trust in both the outputs of LLMs and the models themselves. Moreover, our investigation unveils a significant opposition to the incorporation of fingerprinting features within LLMs. Through extensive data gathering, we gained valuable insights into the prospective applications of LLMs across various sectors, as well as identified potential security vulnerabilities that may arise in the future. Despite inherent limitations, our study contributes substantially to the ongoing efforts aimed at fortifying the safety and resilience of LLMs, thus advancing the discourse on responsible AI deployment and governance. Future research would attain better understanding of future risks by engaging more relevant stakeholders in the field of AI, Ethics, Psychology and Law. By incorporating insights from these stakeholders, model developers would be provided with an ethical

framework, perspectives on effective regulatory measures and a better understanding between LLM features and human psychology aiding in the effort for responsible model development. Furthermore, participants emphasized the necessity for a permission model that empowers them to engage with LLMs with considerations of its limitations and with an understanding of how their personal information will be handled within the model. Incorporating clear consent mechanisms will help enhance trust and facilitate responsible AI usage in society. Also, as demonstrated from the survey responses regarding risks of MLLMs, we believe that efficient watermarking techniques for contents generated by LLMs should be developed. (All data and source code will be made available at: https://github.com/bendjko/CSE-527S-Final-Project.git)

8. Acknowledgments

We thank all our respondents for their thoughtful survey responses. We send a special thanks to Professor Iqbal who helped us steer the direction of the study and provided constructive feedback for both survey design and analysis.

9. References

[1] Particip-ai: A democratic surveying framework for anticipating future AI use cases, harms and benefits: 2024. *https://doi.org/10.48550/arXiv.2403.14791*. Accessed: 2024-05-01.

[2] Public perceptions of gender bias in large language models: Cases of chatgpt and Ernie: 2023. *https://doi.org/10.48550/arXiv.2309.09120*. Accessed: 2024-05-01.

[3] Security and privacy challenges of large language models: A survey: 2024. *https://doi.org/10.48550/arXiv.2402.00888*. Accessed: 2024-05-01.

[4] Ferret: Refer and ground anything anywhere at any granularity: 2023. *https://doi.org/10.48550/arXiv.2310.07704*. Accessed: 2024-05-01.

[5] The World of Generative AI: Deepfakes and large language models: 2024. *https://doi.org/10.48550/arXiv.2402.04373*. Accessed: 2024-05-01.

10. Appendix

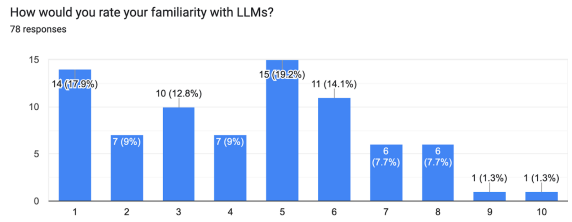

Figure 1.

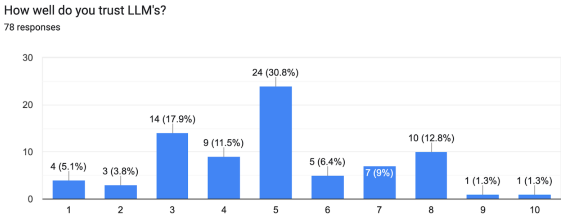

Figure 2.

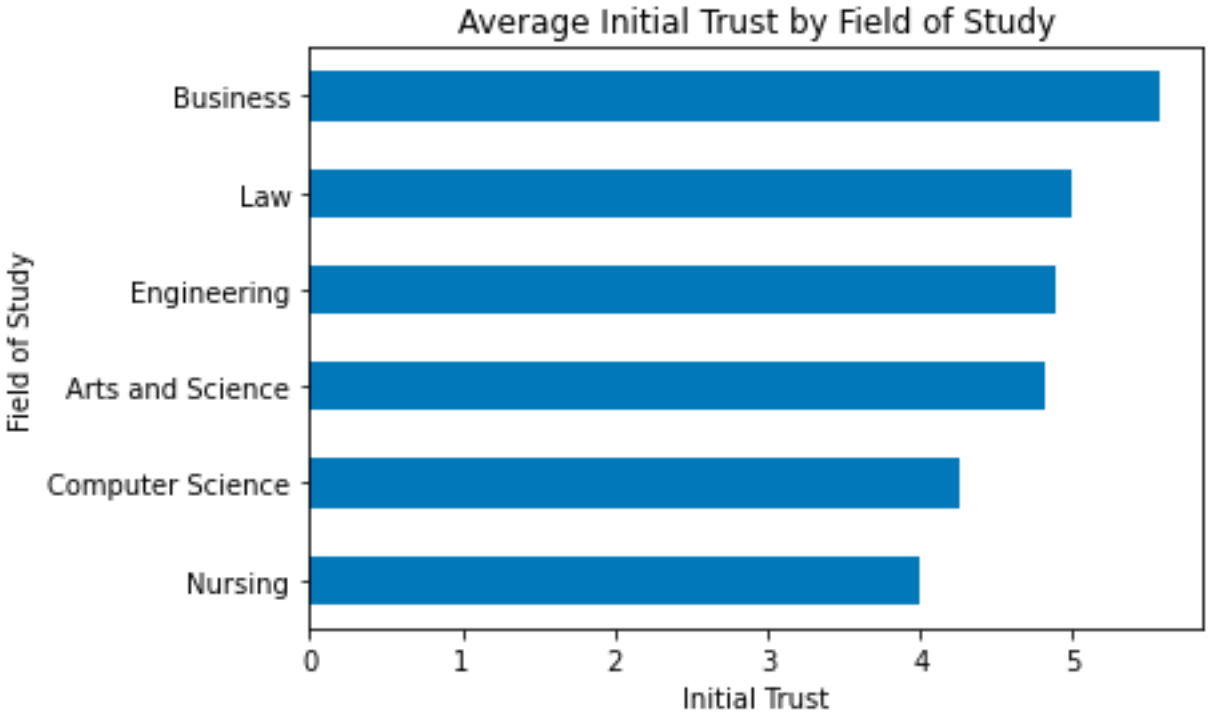

Figure 3.

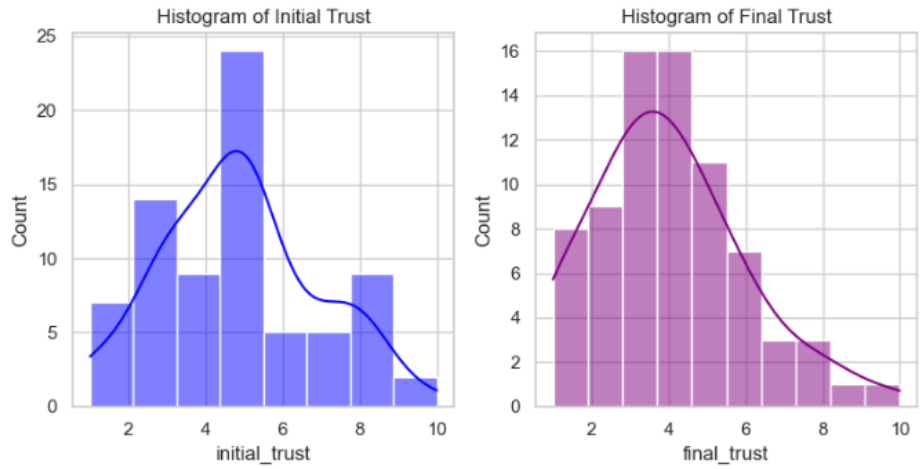

Figure 4.

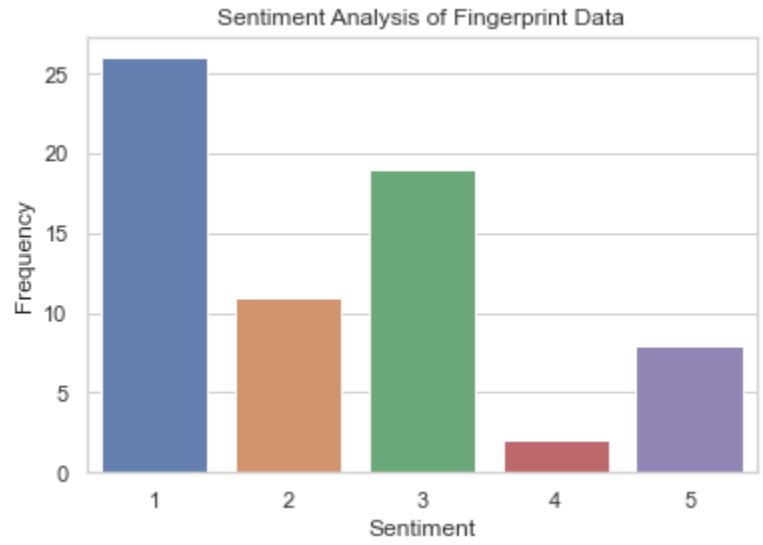

Figure 5.

	Main Category	Frequency	Category
0	Automated Tasks	17	brainstorming, candidate screening, customer service, grammar editing, pattern matching, perform routine tasks, scheduling, summarizations, task summarization
1	Content Generation	23	content creation, content generation, create designs beyond human imagination, creating architectural plans, creating kids bedtime stories, generating animations, generating content, generating lesson plans, generating movies, generating scripts, generating workout plans, language translation, music generation, news article generation, speech writing, writing jokes
2	Education Assistance	4	create lesson plans, homework assistance, study aide
3	Healthcare Assistance	12	analyzing medical scans, diagnose patients, patient diagnosis, patient interaction, patient interaction analysis, phi access, telehealth, virtual therapist
4	Human Resources Tasks	3	interviewing, mitigating bias
5	Legal Assistance	2	legal research
6	Marketing and Advertising	3	personalized advertising, targeted advertising
7	Other	15	all, generating novels, increasing human awareness, mitigate research bias, optimize data, study tool, tasks oraganization, translation, urban planning, virtual reality
8	Research Assistance	9	building financial models and financial research, finding resources, information queries, mapping tumors, research, research assistance, research synthesis, targeted research
9	Social	2	virtual companion, virtual human interactions
10	Software development	27	adaptive programming assistance, advanced code generation, automated writing, creating models, debugging code, generating code, migrating legacy codebase

Figure 6.

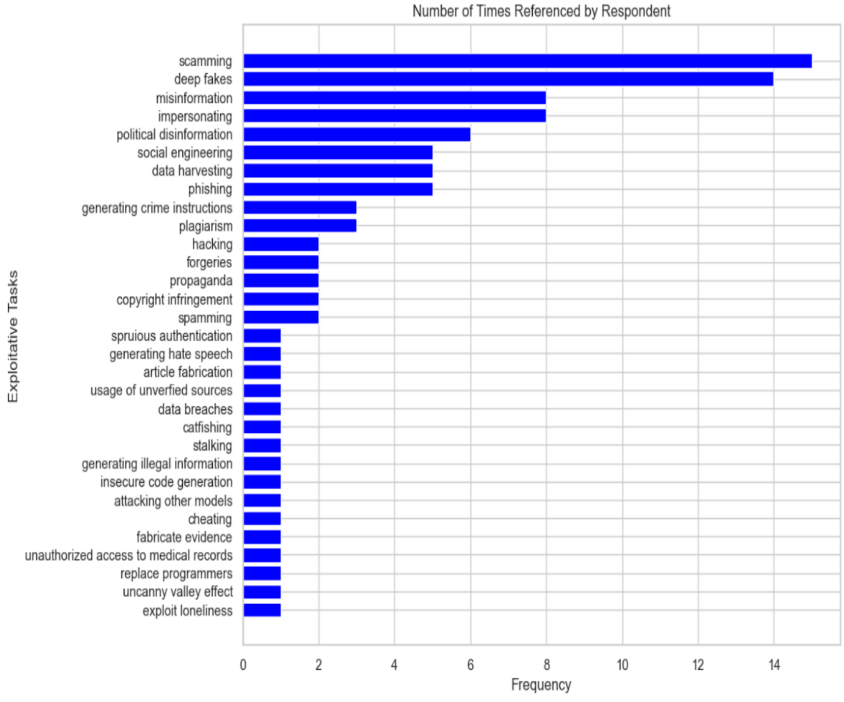

Figure 7.

Main Category	Frequency	Category
Arts and Media	18	arts and entertainment, fashion, journalism, media
Business and Marketing	10	advertisement, customer service, finance, marketing, sales
Crime	2	criminal justice, social engineering
Education	11	education
Finance	2	banking, insurance
Government/Military	11	government, military, politics
Healthcare	13	healthcare
Other	3	manufacturing, social media
Research	2	research
Technology	11	cybersecurity, data science, software development

Figure 8.

	Main Category	Frequency	Category	Percent
0	Data security and privacy	41	compliance with data protection laws, data harvesting, deep fakes, encryption, intellectual property ambiguity, intellectual property theft, misinformation, privacy breaches, selling user data, usage of unverified information	46.6%
1	Education and research	9	academic integrity, bias in research data, decrease in human creativity, fabrication of research data, overreliance on ai in education, overreliance on technology	10.2%
2	Healthcare	7	hipaa violations, patient health information leaked, protecting patient files, unsafe medical advice, untested health care	8.0%
3	Legal and compliance issues	3	impersonation and scams, legal research complexities	3.4%
4	Other	16	biased output, compromised data analysis. misinformation, data poisoning. compromised data analysis, data selling, impersonation, incorrect output, jail breaking, misinformation. patient health information leaked, overreliance on technology for routine tasks, phishing, prompt interjection, scam calls, sensitive data analysis performed public servers, spamming, stalking, usage of unverfied sources	18.2%
5	Programming	12	data poisoning, debugging issues, insecure code generation, jailbreaking, prompt injection, security vulnerabilities in coding, skewed training data, unsecure code generation	13.6%

Figure 9.

	Main Category	Category	Solutions
0	Other	corporate espionage, data leekage, data storage, leveragig llm's to generating password combination, unsecure data trnsmission	restrict llm usage for companies handling sensitive information
1	consent issues	data harvesting, data sale, generation of user profile based on chat history, model retraining using user data, selling data to brokers	regulations for optimizing models using user input, legislation, delete sensitive history
2	data breach	data breach, hacking llm data stores, jailbreaking to reveal proprietary data, prompted injection	encryption for data sent between llms, implement robust llm security measures, anonymized sensitive data storage, localize llm usage
3	invasions of privacy	chat history leakage, exploitation of present visual and audio identity verification mehtods, identity theft, leveraging llms to generate password combinations, subpoena of user history	identity verification before registration, warn users of the risks of inputting sensitive information, restrict llms to non-sensitive tasks, training on pii or being given access to pii for augmentation, training users on llms to ensure safe interactions

Figure 10.

Category	Risks
Individual	false or wrong medical diagnosis, identity theft, unconsented data collection, deep fakes, overuse of AI disrupting patient-doctor relationship, model retraining using user data, generation of user profile based on chat history, leveraging llms to generate password combinations, subpoena of user history, biased medical diagnostics harming minority groups, intellectual property theft, phishing, misinformation, spamming, stalking, incorrect model output, exploitation of visual and audio identity verification methods
Society	loss of human interaction, dependency on LLMs leading to social and cognitive decline, loss of human critical reasoning skills, results becoming unmotivated and decontextualized, flooding journals with fake articles, generation of fake or biased content, misinformation, loss of human interaction, biased medical diagnostics harming minority groups, impersonation and scams
Industry	data breaches, corporate espionage, insecure data transmission, selling user data, loss of jobs due to AI substitution, decrease in strategic and creative business decisions, overreliance on AI leading to less competent employees, proprietary data being stolen, automated vulnerability testing used for hacking, data poisoning, prompt injection, data leakage, security vulnerabilities in coding, insecure code generation, debugging issues, jailbreaking, skewed training data, legal research complexities, disservice to the development of young lawyers, job displacement of support staff, data harvesting, selling user data, hipaa violations, patient health information leaked, untested health care, unsafe medical advice, overreliance on technology for routine tasks

Figure 11.

Figure 12.