Inferring Personally Identifiable Information (PII) from LLM Interactions: Risks, Patterns, and Mitigation Strategies

1st Vineeth Konjeti
Tickle College of Engineering
University of Tennessee, Knoxville
Knoxville, United States
vkonjet1@vols.utk.edu

2nd Julianna Prater
Tickle College of Engineering
University of Tennessee, Knoxville
Knoxville, United States
jprater8@vols.utk.edu

3rd Madina Mirusmanova
Tickle College of Engineering
University of Tennessee, Knoxville
Knoxville, United States
mmirusma@vols.utk.edu

4th Venkatesh Jorige
Tickle College of Engineering
University of Tennessee, Knoxville
Knoxville, United States
jvenkat1@vols.utk.edu

5th Trinity Bissahoyo
Tickle College of Engineering
University of Tennessee, Knoxville
Knoxville, United States
tbissaho@vols.utk.edu

***Abstract*—This project aims to examine the current data handling practices of LLM companies and how that data could be used by LLMs themselves, companies, or bad actors to infer personally identifiable information (PII) about a specific person. Our goal is to raise awareness of how consumers can take steps to ensure their sensitive data remains protected and provide guidance to LLM companies on the best practices for handling data.**

***Index Terms*—AI, artificial intelligence, data, privacy, security, cybersecurity, LLM**

I. Introduction

This project will investigate how different types of data shared with large language models (LLMs) can be used to infer personally identifiable information (PII). We aim to identify which kinds of data are most vulnerable, how much data is needed to infer specific types of PII with varying levels of accuracy, and how model architecture or intended use (e.g., conversational vs. analytical) affects this risk. Our stretch goal is to propose mitigation strategies that AI companies could implement to reduce inadvertent PII exposure. As LLMs are increasingly integrated into daily life, users often share sensitive information without realizing the potential risks. We are seeking to raise awareness of data privacy in LLM interactions and promote best practices for smart usage.

II. Project Approach

This project will draw on a diverse set of data sources to explore how PII can be inferred from interactions with LLMs. We will examine public datasets related to recent cyberattacks, particularly those targeting AI companies, to understand how data vulnerabilities are exploited. Documentation on data scraping and storage practices by major LLM providers such as ChatGPT and Gemini will help us assess what kinds of user data are retained and how they might be used. We will also analyze GPS tracking and social media mining patterns to explore how location data can be inferred or reconstructed. Because LLMs rely on scraping data from the internet for training, one interesting angle of data privacy we'd like to explore is mindfulness of what kind of data a user posts online. Password behavior datasets will be studied to identify common patterns and substitutions that may reveal personal details. In addition, we will incorporate findings from LLM usage logs and behavioral studies, focusing on query types, frequency, and time-of-day patterns that could hint at user identity or location. Finally, we will examine the kinds of questions users typically ask LLMs and how those queries might inadvertently expose sensitive information.

We're planning to approach the subject of data privacy and LLMs from multiple angles and study data from multiple sources. While our main focus will be on the data users enter into LLMs and how LLM companies handle that data, we will also discuss data privacy from a more user-focused perspective. LLMs get their training data by scraping data from the Internet, some of which could include personal data such as social media posts. From there, LLMs could extract a lot of data about a particular person from all the data they generate just from being on the internet and upload to the internet. Not only could it jeopardize the anonymity of the user, it could also jeopardize a person's safety if another person were to ask an LLM about them.

III. Expected Outcome

By the end of the project, we expect to produce a curated reference list of all datasets and academic papers consulted, along with a Jupyter notebook that documents our data cleaning, analysis, and visualizations. Our findings will include a breakdown of which types of user data are most vulnerable to PII inference, how much data is typically required to make accurate inferences, and how different model architectures or use cases affect the risk level. We also aim to propose a set of mitigation strategies that users and AI companies can adopt to reduce the likelihood of inadvertent data exposure. These strategies will be grounded in both technical analysis and ethical considerations, offering practical recommendations for smarter LLM data safety practices.

IV. Constraints and Considerations

In conducting this research, we will remain mindful of several key constraints. Dataset size and availability may limit the scope of our analysis, particularly when working with proprietary or restricted data. We will prioritize publicly accessible sources and supplement them with secondary materials such as peer-reviewed papers and technical documentation. Our focus will remain on LLMs specifically, rather than broader AI systems, to ensure the relevance and precision of our findings. Ethical implications will be considered throughout the project, especially in relation to data mining, the potential misuse of inferred information, and potentially sensitive information that we may find in datasets. We aim to balance comprehensive research and analysis with responsible research practices, ensuring that our work contributes meaningfully to the conversation around AI and data privacy.

V. Project Timeline

For this project, we aim to solidify our dataset and figure out our analysis methods and visualizations toward the beginning of October. Between mid-October and early November, we'd like to

- **October:** Solidify datasets + dfigure out analysis methods and visualizations
- **November:** Run analysis + document findings + draw conclusions and comparisons among our own and each other's data.
- **Late November:** Prepare for presentation + finish final draft of paper

VI. Team Responsibilities

To develop a comprehensive dataset, each member has a specific area of interest within PII data to research. These datatypes include LLM user habits, LLM data storage practices, common password practices, and GPS and social media data and usage habits. Members are responsible for running data analysis on their given datatype, with at least one common data analysis method among the group to ensure data comparison is conducive. Based on the findings from the research and data testing, each member draws up a discussion on what was observed in their specific category. Coming together, each member's findings will be compared with each other to develop a conclusion and future research on PII risk produced by LLMs. Communication throughout this project is to remain consistent among members, meeting once a week to discuss progress and develop the project.

References

[1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] K. Elissa, "Title of paper if known," unpublished.

[5] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.