ETH zürich

Master Thesis: Inferential privacy impact of LLM-based chatbot usage

Dr. Noé Zufferey

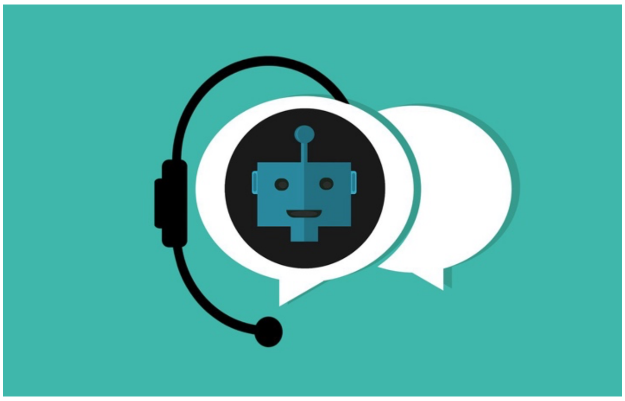
Security, Privacy and Society
D-GESS, ETHZ

Security
Privacy &
Society

Background

The performance of artificial intelligence, in particular, generative artificial intelligence (GenAI) has rapidly increased over the last few years and recently reached tremendous efficiency, especially regarding text and image generation. These models are trained with many different types of data, including personal data, to achieve many different tasks.

In this project, we aim to use an inferential privacy approach in order to quantify the privacy of LLM-based tools users, and more specifically users of LLM-based chatbots like ChatGPT.



To what extent can an LLM-based chatbot infer an individual's private information from their interaction?

In this research project, we aim to focus on the assessment of the threats to privacy that can be raised by the use of large-language-model (LLM) based chatbots. The goal of this research project is to quantify how such tools can be a threat to users' privacy.

In this project, we aim to quantity LLM-based chatbot users' privacy by using an inferential privacy approach. Such an approach asks the researchers to embody the adversary by:
 a literature of existing work in that area

- listing and collecting data
- conducting an inference attack
- evaluating how successful the attack is

Therefore, the main investigator will be in charge of developing and fine-tuning an LLM model in order to evaluate to what extent such tools may be used to infer private users' information from discussion records.

This project requires programming skills in Python and previous basic knowledge of Neural Networks and/or Generative Machine learning.

The Security, Privacy and Society Group

The group conducts research at the intersection of humans and technology with a focus on human-computer interaction and the human aspects of IT security and privacy. Using interdisciplinary approaches, we aim to tackle current security-related challenges and to develop human-centered solutions.

ETH Zürich
Dr. Noé Zufferey
STB G18.2
Stampfenbachstrasse 69
8092 Zürich

noe.zufferey@gess.ethz.ch
www.spg.ethz.ch

D GESS