DataGuardian: AI-Powered Multi-Modal Visual and Textual Data Anonymization System

Karamjeet Singh Gulati, Jason Yoo, Nikita Emberi
ksgulati@ucdavis.edu, ysyoo@ucdavis.edu, nemberi@ucdavis.edu

Question

In today's digital age, there is an increasing need to protect personally identifiable information (PII) present in multi-modal data, such as images and text, from exposure. The central question our project aims to address is:

How can an AI-powered system effectively anonymize both visual and textual data simultaneously, while preserving the utility of the data for research and analysis purposes?

This question will guide the development of our research and proposed solution for ensuring privacy in data usage.

Problem Statement

The volume of sensitive data generated and shared digitally continues to grow exponentially, often containing personally identifiable information (PII) across both visual and textual formats. These include images with visible faces and documents with sensitive details like names, addresses, and financial information. Current anonymization solutions tend to focus on either visual or textual data, leaving a critical gap in the comprehensive protection of multi-modal data. Our project seeks to explore how AI can be leveraged to close this gap, addressing the following sub-questions:

- What challenges arise in anonymizing multi-modal data?
- How can anonymization maintain data utility while ensuring privacy?

Proposed Solution

DataGuardian will address the challenge by developing an AI-powered system capable of anonymizing both visual and textual data. We propose the following features for the system:

Multi-Modal AI Analysis:

- Use OpenAI's GPT-4 with vision capabilities to analyze both images and textual input simultaneously.
- Perform cross-modal analysis to better identify sensitive information across formats.

Adaptive Data Anonymization:

- Develop custom recognizers for detecting various types of sensitive information in both text and images.
- Apply tailored anonymization techniques (e.g., blurring faces, redacting text) based on the specific data type.

Interactive User Interface:

- Build a Gradio-based interface for real-time data input and visualization of anonymized outputs.
- Create an intuitive dashboard to present analysis results and allow users to adjust settings if necessary.

Privacy-First Architecture:

- Ensure compliance with data privacy standards like GDPR and HIPAA.
- Implement secure data handling, storage, and anonymization practices.

The system's workflow involves capturing or inputting data, analyzing it using AI to identify sensitive information, applying appropriate anonymization techniques, and presenting the anonymized output to the user. This process is performed in real-time, allowing for immediate feedback and adjustment.

By offering a unified solution for both visual and textual data anonymization, DataGuardian provides a more comprehensive approach to data privacy. It not only protects individual privacy but also enables organizations to safely utilize and share valuable data for research and analysis purposes.