Data Anonymization for LLM-based Automotive Requirements Processing

Description

This thesis will focus on developing methods to protect privacy while using Large Language Models (LLMs) in automotive requirements processing. As LLMs process sensitive data, such as customer information and design specifications, ensuring that this data remains private is crucial. The research will explore anonymization techniques like data masking and pseudonymization, which hide sensitive details while keeping the data useful for LLM tasks. It will also focus on ensuring that anonymized data retains its meaning and effectiveness in the context of automotive requirements processing.

For application please send me an email with title "Master Thesis Application: Data Anonymization LLM ". Please also attach your resume and transcript of records in the email. A motivation letter is NOT required.

Tasks

- Literature review on data privacy enhancement for LLMs, e.g., [1].
- Literature review on regulations about general data privacy, automotive-related sensitive information, e.g., GDPR.
- Create automated tools to anonymize LLM input.
- Evaluation and benchmark of the proposed approach.

References

[1] Ahmed Frikha, Nassim Walha, Krishna Kanth Nakka, Ricardo Mendes, Xue Jiang, and Xuebing Zhou. Incognitext: Privacy-enhancing conditional text anonymization via llm-based private attribute randomization, 2025.

Technische Universität München

TUM School of Computation, Information and Technology

Lehrstuhl für Robotik, Künstliche Intelligenz und Echtzeitsysteme

---

Supervisor:
Prof. Dr.-Ing. Alois Knoll

Advisor:
Fengjunjie PAN, M.Sc.

Research project:
MANNHEIM-CeCaS

Type:
MA

Research area:
Generative AI, Large Language Models, Data privacy

Programming language:
Python

Required skills:
LLM deployment, Automation Tools, Programming language

Language:
English

For more information please contact us:

E-Mail: f.pan@tum.de

Internet: www.ce.cit.tum.de/air