RAG with Differential Privacy

Nicolas Grislain

December 5, 2024

Abstract

This is a document abstract.

Introduction

Related Work

A reference (Abadi et al. 2016)

Some solutions are based on privacy preserving synthetic data generation: (Zeng et al. 2024)

(Ponomareva et al. 2023)

(Lebensold et al. 2024)

(Lin et al. 2024)

(Xie et al. 2024)

(Tang et al. 2024)

(Wu et al. 2023)

(Hong et al. 2024)

DP-RAG

Overview

Privacy Unit Preserving Document Retrieval

Differentially Private In-Context Learning

Evaluation

Conclusion

Abadi, Martin, Andy Chu, Ian Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. 2016. "Deep Learning with Differential Privacy." In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. CCS'16. ACM. https://doi.org/10.1145/2976749.2978318.

Hong, Junyuan, Jiachen T. Wang, Chenhui Zhang, Zhangheng Li, Bo Li, and Zhangyang Wang. 2024. "DP-OPT: Make Large Language Model Your Privacy-Preserving Prompt Engineer." https://arxiv.org/abs/2312.03724.

Lebensold, Jonathan, Maziar Sanjabi, Pietro Astolfi, Adriana Romero-Soriano, Kamalika Chaudhuri, Mike Rabbat, and Chuan Guo. 2024. "DP-RDM: Adapting Diffusion Models to Private Domains Without Fine-Tuning." https://arxiv.org/abs/2403.14421.

Lin, Zinan, Sivakanth Gopi, Janardhan Kulkarni, Harsha Nori, and Sergey Yekhanin. 2024. "Differentially Private Synthetic Data via Foundation Model APIs 1: Images." https://arxiv.org/abs/2305.15560.

Ponomareva, Natalia, Hussein Hazimeh, Alex Kurakin, Zheng Xu, Carson Denison, H. Brendan McMahan, Sergei Vassilvitskii, Steve Chien, and Abhradeep Guha Thakurta. 2023. "How to DP-Fy ML: A Practical Guide to Machine Learning with Differential Privacy." *Journal of Artificial Intelligence Research* 77 (July): 1113–1201. https://doi.org/10.1613/jair.1.14649.

Tang, Xinyu, Richard Shin, Huseyin A. Inan, Andre Manoel, Fatemehsadat Mireshghallah, Zinan Lin, Sivakanth Gopi, Janardhan Kulkarni, and Robert Sim. 2024. "Privacy-Preserving in-Context Learning with Differentially Private Few-Shot Generation." https://arxiv.org/abs/2309.11765.

Wu, Tong, Ashwinee Panda, Jiachen T. Wang, and Prateek Mittal. 2023. "Privacy-Preserving in-Context Learning for Large Language Models." https://arxiv.org/abs/2305.01639.

Xie, Chulin, Zinan Lin, Arturs Backurs, Sivakanth Gopi, Da Yu, Huseyin A Inan, Harsha Nori, et al. 2024. "Differentially Private Synthetic Data via Foundation Model APIs 2: Text." https://arxiv.org/abs/2403.01749.

Zeng, Shenglai, Jiankun Zhang, Pengfei He, Jie Ren, Tianqi Zheng, Hanqing Lu, Han Xu, Hui Liu, Yue Xing, and Jiliang Tang. 2024. "Mitigating the Privacy Issues in Retrieval-Augmented Generation (RAG) via Pure Synthetic Data." https://arxiv.org/abs/2406.14773.