Shenglai Zeng

https://slz-ai.github.io/ shenglaizeng@gmail.com (+1)517-9747616

RESEARCH INTEREST

I am a second-year PhD student at Michigan State University advised by Professor Jiliang Tang. My current research interests are mainly about Trustworthy AI, large language models(LLMs) and Information retrieval (IR). I also worked on federated learning for years.I have won the Best Paper Award of IEEE Transactions on Cloud Computing, 2023.

EDUCATION

Michigan State University **East Lansing, U.S**
DSE Lab/PhD students in Computer Science and Engineering *Sept 2023-Present*
Advisor: Jiliang Tang
Lab: Data Science and Engineering Lab
Research Direction: Trustworthy AI, Large language models(LLMs)

University of Electronic Science and Technology of China **Chengdu, China**
Yingcai Honor School/B.Sc in Computer Science and Engineering *Sept 2019-Present*
CGPA: 3.98/4.00
Weighted Average: 93.97/100(1st among 100 students)
Honors: The Most Outstanding Students Award of UESTC

PREPRINTS

- **Shenglai Zeng**, Jiankun Zhang, Pengfei He, Jie Ren, Tianqi Zheng, Hanqing Lu, Han Xu, Hui Liu, Yue Xing, Jiliang Tang
  Mitigating the privacy issues in retrieval-augmented generation (rag) via pure synthetic data
  Submitted to ACL ARR
- Jie Ren, Han Xu, Pengfei He, Yingqian Cui, **Shenglai Zeng**, Jiankun Zhang, Hongzhi Wen, Jiayuan Ding, Hui Liu, Yi Chang, Jiliang Tang
  Copyright Protection in Generative AI: A Technical Perspective
  Pre-print

PUBLICATIONS

- **Shenglai Zeng**, Jiankun Zhang, Bingheng Li, Yuping Lin, Tianqi Zheng, Dante Everaert, Hanqing Lu, Hui Liu, Yue Xing, Monica Xiao Cheng, Jiliang Tang
  Towards Knowledge Checking in Retrieval-augmented Generation: A Representation Perspective
  NAACL-2025-main(Oral)
- Jie Ren, Kangrui Chen, Yingqian Cui, **Shenglai Zeng**, Hui Liu, Yue Xing, Jiliang Tang, Lingjuan Lyu
  Six-cd: Benchmarking concept removals for benign text-to-image diffusion models
  CVPR-2025
- Pengfei He, Yue Xing, Han Xu, Jie Ren, Yingqian Cui, **Shenglai Zeng**, Jiliang Tang, Makoto Yamada, Mohammad Sabokrou
  Stealthy Backdoor Attack via Confidence-driven Sampling
  TMLR
- Jie Ren, Yaxin Li, **Shenglai Zeng**, Han Xu, Lingjuan Lyu, Yue Xing, Jiliang Tang

Unveiling and mitigating memorization in text-to-image diffusion models through cross attention
ECCV-2024

- Han Xu,Jie Ren,Pengfei He, Yingqian Cui, **Shenglai Zeng**, Hui Liu, Jiliang Tang, Amy Liu
  On the Generalization of Training-based ChatGPT Detection Methods
  EMNLP-2024-Findings
- **Shenglai Zeng***, Yaxin Li*, Jie Ren, Yiding Liu, Han Xu, Pengfei He, Yue Xing, Shuaiqiang Wang, Jiliang Tang, Dawei Yin
  Exploring Memorization in Fine-tuned Language Models
  ACL-2024
- **Shenglai Zeng***, Jiankun Zhang*, Pengfei He, Yue Xing, Yiding Liu, Han Xu, Jie Ren, Shuaiqiang Wang, Dawei Yin, Yi Chang, Jiliang Tang
  The Good and The Bad: Exploring Privacy Issues in Retrieval-Augmented Generation (RAG)
  ACL-2024-findings
- Pengfei He, Han Xu, Jie Ren, Yingqian Cui, **Shenglai Zeng**, Hui Liu, Charu Aggarwal, Jiliang Tang
  Sharpness-aware Data Poisoning Attack
  ICLR-2024 Spotlight
- Juanhui Li, Harry Shomer, Haitao Mao, **Shenglai Zeng**, Yao Ma, Neil Shah, Jiliang Tang, Dawei Yin
  Evaluating graph neural networks for link prediction: Current pitfalls and new benchmarking
  NIPS-2023 Benchmark
- **Shenglai Zeng**, Zonghang Li, Hongfang Yu, Zhihao Zhang, Long Luo,Bo Li, Dusit Niyato
  HFedMS: Heterogeneous Federated Learning with Memorable Data Semantics in Industrial Metaverse
  IEEE Transactions on Cloud Computing, 2023 Best Paper
- **Shenglai Zeng**, Zonghang Li, Hongfang Yu, Yihong He, Zenglin Xu, Dusit Niyato, Han Yu
  Heterogeneous Federated Learning via Grouped Sequential-to-Parallel Training
  International Conference on Database Systems for Advanced Applications (DASFAA-2022)
- Jiaqi Wang*, **Shenglai Zeng***, Zewei Long, Yaqing Wang, Houping Xiao, Fenglong Ma
  Knowledge-Enhanced Semi-Supervised Federated Learning for Aggregating Heterogeneous Lightweight Clients in IoT
  SDM-2023

Patent

- **Shenglai Zeng**, Zonghang Li, Yihong He, Xun Zhang, Hongfang Yu, Gang Sun
  "A Hierarchical User Training Management Architecture and Training Strategy for Non-i.i.d Data" Chinese patent

RESEARCH EXPERIENCE

DSE Lab, Michigan State University **Lansing, U.S**
Research Assistant/Research on Trustworthy AI and LLM-safety *Sept 2023 - Present*

-Advisor: Professor Jiliang Tang

- Identify and mitigate the real privacy issues of LLMs.
- Diverse attack/defense techniques on LLM systems.
- Deeper understanding of underlying mechnism behind LLMs, especially knowledge extraction perspective.
- Leaverage LLMs to enhance/empower challanging applications and tasks.

Search Science Team, Amazon **CA, USA**
Research Intern/Research on knowledge-checking in RAG *May 2024 - Present*

-Mentor: Tianqi Zheng, Dante Everaert
-Manager: Hanqing Lu
- Investigating robustness issues of RAG.
- Utilize LLMs' internal behavior to conduct knowledge checking in RAG.
- Enhance the performance by representation-based context filtering.

Search Science Team, Baidu.Inc **Beijing, China**
Research Intern/Research on the memorization of LLM *May 2023 - May 2024*

-Mentor: Dr. Yiding Liu and Dr. Dawei Yin
- Investigating the memorization behavior and privacy implications of fine-tuned LLMs.
  One conference Paper submitted to ICLR 2023(First author)
- Currently worked on privacy risks of Retrieval LMs and AI-agents.

Intelligent Networking and Applications Research Center, UESTC **Chengdu, China**
Research Assistant/Research on the Optimization of Federated Learning *Sept 2020 - Jun 2023*

-Mentor: Professor Hongfang Yu and Dr.Zonghang Li
- Proposed a novel idea of Sequential-to-Parallel training in FL.
  One conference Paper accepted by DASFAA 2022(First author)
- Investigated the application of FL in Industrial Metaverse.
  One journal paper accepted by IEEE TCC(First author).

The Pennsylvania State University **Pennsylvania, USA**
Online Intern/Research on Semi-supervised Federated Learning *Jun 2021 - Jun 2022*

-Mentor: Professor Fenglong Ma
- Implemented a semi-supervised federated learning system combined with novel personalized punning and structure-aware collaborative distillation techniques.
  Paper accepted by to SDM 2022(Co-First Author)
- Currently focusing on FL with different model structures.

University of British Columbia **Vancouver, Canada**
Summer Intern/Federated Data Evaluation with Unlearning *Jun 2022 - Sept 2022*

-Mentor: Professor Xiaoxiao Li
- Try to use the concept of cooperative game to evaluate the importance of data of participating users in federated learning.
- Try to accelerate the evaluation process and straggler problem using federated unlearning.

The University of Chicago **Chicago, USA**
Online Intern/Research on the IOT & Sensing Security *Mar 2020 - Mar 2021*

-Mentor: Shinan Liu(PhD candidate)
- Tried to use audio data collected by microphone to reconstruct user's state of motion during recording time.
- Proposed a mathematical-physical model to explain the correlation between different sensors' responses to motion.

KEY SKILLS

Programming Language	Python, C, C++, Java,Matlab
Research Tool	Latex, Overleaf
AI Framework	Pytorch, Mxnet,Tensorflow
Network	Cisco Certified Network Associate(CCNA)

SELECTED ACADEMIC PROJECTS

Exploring Privacy Issues in Retrieval-Augmented Generation (RAG) *Oct 2023- Feb 2024*

- We've uncovered two pivotal aspects: (1). Privacy challenges within RAG's own data (2). RAG's potential to safeguard training data
- **Data Leak Quantified:** RAG systems can leak private retrieval data, with our study showing about 50% of sensitive retrieval data being output.
- **Mitigation Efforts:** We've explored naive defenses such as summarization and retrieval thresholds. These methods help mitigate risks but don't completely resolve the issue, indicating the gravity of privacy risks in RAG.
- **Training Data Safeguard:** RAG shows promise in protecting training data, offering a strategy to bolster privacy in AI systems.
- Our code is available at https://github.com/phycholosogy/RAG-privacy

Exploring Memorization in Fine-tuned Language Models *May 2023- Oct 2023*

- Extensively studied the memorization effect of LLMs during the fine-tuning stage across different tasks.
- **Fine-tuning Risks:** Utilizing copyrighted or private data in fine-tuning poses privacy/IP risks.
- **Task Disparity:** Summarization & Dialogue show high memorization, while QA and Classification are lower.
- **Task-specific Scaling:** For high memorization tasks, memorization increases with larger models. Conversely, for low memorization tasks, increasing model size has little impact on memorization.
- **Attention's Role:** High memorization tasks have uniform, sparse attention patterns. We unravel the nuances between attention & memorization.
- **Solution in Sight:** Multi-task fine-tuning buffers against high memorization vs single-task techniques.

Federated Learning Framework Design in Industrial Metaverse *Feb 2022 – Present*

- **Background**: This work mainly focuses on the non-i.i.d streaming data collected by distributed edge devices in Industrial Metaverse.(e.g. OCR applications in industrial parks.)
- **Challenge**: data heterogeneity/limited band width/ catastrophic forgetting towards streaming data
- A dynamic FL training paradigm designed for rapidly changing streaming data while eliminating data heterogeneity.
- A knowledge maintained online learning method for FL to prevent catastrophic forgetting
- Task decomposition in federated learning to reduce communication pressure on edge networks
- This work is accepted by IEEE Transactions on Cloud Computing.

Semi-Supervised Federated Learning for IoT *Jun 2021 – Sept 2022*

- This work aims to develop an applicable federated learning mechanism for IoT devices in semi-supervised settings, considering personalization and communication efficiency simultaneously.
- We introduced neural network pruning techniques into semi-supervised federated learning and a novel structure-aware collaborative distillation approach which can aggregate models with different structures.
- This work is accepted by SDM 2023.

AWARDS AND ACHIEVEMENTS

- **Best Paper Award, IEEE Transactions on Cloud Computing**
- **The Most Outstanding Students Award of UESTC** (Highest honor in UESTC, Only10 students are awarded)
- **National Scholarship** in the session of 2019-2020.(Highest honor of undergraduate student)
- WAC Scholarship in the session of 2020-2021.(Only 10 undergraduate students in UESTC are awarded each year)
- 1st Outstanding Academic Scholarship in 2020,2021,and 2022.(Top 5 % students)

Services

- **Program Committee or Reviewer:** IEEE TKDE, IoT-J, IEEE Trans on Information Forensics Security, MICCAI Workshop on Distributed, Collaborative and Federated Learning (DeCaF-2022), The ACM Transactions on Information Systems (TOIS),IEEE Transactions on Vehicular Technology(TVT),IEEE TKDD, ACL-ARR
- Reviews 10+ papers.