Yixiong Fang

(+86) 131 6210 8292 | kfangyixiong@gmail.com | https://kigb.github.io/

Education

Carnegie Mellon University, Language Technologies Institute
Master of Science in *Intelligent Information Systems* Aug. 2025 – May. 2027 (Incoming)

Shanghai Jiao Tong University (SJTU) Shanghai, China
Bachelor of Engineering in *Software Engineering* Aug. 2021 – Jun. 2025 (Expected)
GPA: 3.9/4.0

Related Courses： Computer Graphics, Computer System Engineering, Computer Vision, Machine Learning, VR/AR & Game Design and Development

Publications

[1] LastingBench: Defend Benchmarks Against Knowledge Leakage. **Yixiong Fang***, Tianran Sun*, Yuling Shi, Min Wang, Xiaodong Gu.

[2] AttentionRAG: Attention-Guided Context Pruning in Retrieval-Augmented Generation. **Yixiong Fang**, Tianran Sun, Yuling Shi, Xiaodong Gu.

[3] Yupei Li*, **Yixiong Fang***, Lucia Specia, Xiaodong Gu, Björn Schuller "Enhancing Speech Large Language Models for Deepfake Speech and Audio Detection through Feature Dropout and Dropin"

[4] **Yixiong Fang***, Ziran Yang*, Zhaorun Chen, Zhuokai Zhao, Jiawei Zhou "From Uncertainty to Trust: Enhancing Reliability in Vision-Language Models with Uncertainty-Guided Selective Decoding"

[5] Yalan Lin, Chengcheng Wan, **Yixiong Fang**, Xiaodong Gu. "CodeCipher: Learning to Obfuscate Source Code Against LLMs"

[6] **Yixiong Fang**, Weixi Yang "Efficient Computation of Eigenvalues in Diffusion Maps: A multi-strategy Approach", 2024 7th International Conference on Signal Processing and Machine Learning

Skills

Programming Languages: C/C++, python, pytorch, Go, Java, javascript, C#, Rust, Coq, MATLAB
Open Source Projects: OceanBase 8.5k stars, MathTranslate 1.1k stars
Languages: Mandarin (native), English (fluent TOEFL 110 GRE 327)

Research Experience

Detecting and Reinforcing Long Context Benchmarks from Leakage **SJTU**
Supervised by Prof. Xiaodong Gu March. 2025 - May. 2025

- Design and implement an disturbance based leakage detection method; systematically reveal the large scale knowledge leakage in long context benchmarks(especially HotpotQA) among various models.
- Propose a pipeline that first detect the leakage and then restore with critical evidence finding and counterfactual rewriting to reinforce the leaked benchmarks without destroying original intent.
- Evaluating models on reconstructed benchmarks and demonstrate the notable performance drops in certain models, which also align with the results of leakage detection(like Qwen3).

Embedding-Level Code Perturbation for Enhanced Security in Large Language Models (LLMs) **SJTU**
Supervised by Prof. XiaoDong Gu Jul. 2024 – Sep.2024

- Design, implement and analysis of the perturbation technique at the embedding level within LLMs, which transformed input tokens into unreadable codes and enhanced security applications.
- Conducted extensive testing on models and migrate the method cross models, validating the effectiveness of the perturbation method.

Mitigating Hallucination in Large Visual Language Models **Stony Brook University**
Supervised by Prof. Jiawei Zhou Apr. 2024 - Present

- Utilize epistemic uncertainty to judge the informative importance of tokens; project visual tokens into language spaces for interpretation; dynamically mask visual tokens for better generation; introduce majority voting strategy in token-level generation.

- Developed evaluation metrics, tested on CHAIR, THRONE and MMBench benchmark and validate its performance in all metrics across models

AttentionRAG: LLM Prompt Compression with Attention **SJTU**
Supervised by Prof. Xiaodong Gu Oct. 2024 – Jan. 2025

- Design question decomposition and answering format for attention key token recognition following a next-token-prediction paradigm; utilizing attention feature of the predicted token for prompt compression
- Conduct experiments on Babilong, LongBench and etc. dataset across models, achieving SOTA results.
- Validate the method's performance is stable with smaller(like 7B compared to 70B) and quantized(like int4) models, providing adaptability in real world scenes.

Deepfake Audio Detection using LLMs Imperial College London
Supervised by Prof. Björn Schuller Dec. 2024 - Present

- Explore the potential of detecting deepfake audio with multimodal LLMs.
- Propose a feature drop-in method and audio level dropout method during training stage, demonstrating an accuracy increase for 20%.

VR-based High-Realism Recovery System **SJTU**
Supervised by Prof. Xubo Yang Apr. 2024 - Present

- Design and develop a high-realism VR system for recovery using Unity
- Collect data of real lake scene and construct virtual lake scene using 3D-Gaussian; Adapt to the interaction between physical equipment and virtual water environments: pairing a rowing machine with a VR kayaking
- Designed mini-games tailored to patients' needs: navigating a kayak toward a marker with the goal of achieving the shortest path, to assist in medical analysis in the impact of cognitive training during recovery

Work Experience

Emagen AI, Tech Leader Present

- Lead the CyberAgent product development(including memory, action, tools, agency parts), manage the developing team of 10+ members

Ant Group Co., Ltd., OceanBase SQL OPTIMIZER TEAM intern Jul. 2024 - Aug. 2024

- Assist open source oceanbase work of OceanBase
- **Enhanced Performance and Efficiency:** Boosted UTF-8 validation speed by 20x-50x in high-production environments, optimized ASCII case batch computations and using SIMD instructions
- Expanded OceanBase's international character encoding support by implementing four East Asian charsets (e.g., ujis), enabling robust data processing for global users and markets
- **User Experience Enhancement:** Developed a virtual table feature for external table errors, allowing users to identify and resolve issues faster during external table construction

Step AI, Agent Development Intern Feb. 2024 - Jun. 2024

- Built backend systems using trpc-go to support the development of the RAG system for Step Chat, integrating PDF parsing, vector databases, and index building for enhanced AI-driven responses.
- Developed tools in agent systems, enabling real-time access to news, weather, and other external data through API integrations with LLM
- Led backend development for the OpenAPI platform, collaborating with product and testing teams to deploy features in test and production environments. Managed virtual account for enterprise clients, data filtering, and automated synchronization to Feishu

Teaching Experience

SJTU Xuesen Challenge Program: High School Hybrid STEM Learning(2025 Winter)
Teaching Assistant for AI1102: Intel's Cutting-Edge AI Algorithms and Practices (2025 Spring)
Teaching Assistant for Mthread GPU Open Class(2025 Spring)