Hao Li

Personal Homepage   18th.leolee@gmail.com   Google Scholar   Github

EDUCATION

Washington University in St. Louis
Ph.D., computer science — *Aug 2025 – Present*
Advised by Prof. Ning Zhang

University of Electronic Science and Technology of China
Master, computer science — *Sep 2021 – Jul 2024*
Co-advised by Prof. Jingkuan Song, Prof. Lianli Gao

Northeast Forestry University
B.S., computer science — *Sep 2017 – Jul 2021*

RESEARCH EXPERIENCE

Washington University in St. Louis — *Aug 2025 – Present*
Superviser: Prof. Zhang Ning

University of Wisconsin–Madison, SaFoLab — *Jul 2024 – Present*
Superviser: Prof. Chaowei Xiao

National University of Singapore, NExT++ — *Jul 2023 – Jul 2024*
Superviser: Prof. Tat-Seng Chua, **Mentor:** Dr. An Zhang

University of Electronic Science and Technology of China, CFM — *Sep 2021 – Jun 2024*
Co-Supervisors: Prof. Jingkuan Song, Prof. Lianli Gao

RESEARCH INTERESTS

LLM Security, Trustworthy Learning, Agentic System

PUBLICATIONS

- **Hao Li**, Jingkuan Song, Lianli Gao, Pengpeng Zeng, Haonan Zhang, Gongfu Li. "A Differentiable Semantic Metric Approximation in Probabilistic Embedding for Cross-Modal Retrieval". ***NeurIPS***, 2022. pdf
- **Hao Li**, Jingkuan Song, Lianli Gao, Xiaosu Zhu, Heng Tao Shen. "Prototype-based Aleatoric Uncertainty Quantification for Cross-modal Retrieval". ***NeurIPS***, 2023. pdf
- Xu Zhang*, **Hao Li***, Mang Ye. "Negative Pre-aware for Noisy Cross-modal Matching". ***AAAI***, 2024.
- **Hao Li***, Chenghao Yang*, An Zhang, Yang Deng, Xiang Wang, Chua Tat-seng. "Hello Again! LLM-powered Personalized Agent for Long-term Dialogue". ***NAACL***, 2025. pdf
- **Hao Li***, Xiaogeng Liu*, Ning Zhang, Chaowei Xiao. "PIGuard: Prompt Injection Guardrail via Mitigating Overdefense for Free". ***ACL***, 2025. pdf
- **Hao Li**, Xiaogeng Liu, Hung-Chun Chiu, Dianqi Li, Ning Zhang, Chaowei Xiao. "DRIFT: Dynamic Rule-Based Defense with Injection Isolation for Securing LLM Agents". ***NeurIPS***, 2025. pdf

PREPRINTS

- **Hao Li***, Jiayang Gu*, Jingkuan Song, An Zhang, Lianli Gao. "One-step Noisy Label Mitigation". ***arXiv***, 2024. pdf
- **Xu Zhang**, Hao Li, Zhichao Lu. "CrossGuard: Safeguarding MLLMs against Joint-Modal Implicit Malicious Attacks". ***arXiv***, 2025. pdf

PROFESSIONAL SERVICE

- Area Chair of ARR 2025 Oct, 2026 Jan.
- Reviewer of CVPR 24.
- Reviewer of ECCV 24.
- Reviewer of ICML 24, 25.
- Reviewer of NeurIPS 24, 25.

- Reviewer of ICLR 25, 26.
- Reviewer of WWW 24.
- Reviewer of AAAI 24, 25, 26.