Li Tianlin

Office mailing address: 50 NANYANG AVENUE Block N4-B2c-06
Email: tianlin001@e.ntu.edu.sg
Contact Number: 89415656

Current Position

Fourth-year Ph.D. student supported by **AISG PhD Fellowship**, supervised by Prof. Liu Yang.

Employment History

Jul. 2019–Jul. 2020
Research Assistant, supervised by Prof. Xianglong Liu, Data Science Group, Beihang University.
Researched on interpreting adversarial vulnerability for deep models in neuron level and designed new defense algorithms.

Mar. 2019–Jun. 2019
Research Assistant, supervised by Prof. Quanshi Zhang, John Hopcroft Center, Shanghai Jiao Tong University.
Researched on interpretability of neural networks, especially on knowledge consistency between pre-trained deep neural networks.

Academic Qualifications

Aug. 2020–Now
Ph.D. Student, Nanyang Technological University
Researching on trustworthy AI.

Sept. 2016–Jan. 2019
M.Eng., Beihang University
Researched on formal verification of MIPS CPU.

Sept. 2012–July. 2016
B.Eng., Beihang University

Author Publications (*: co-first author, #:corresponding author)

Preprint

Purifying Large Language Models by Ensembling a Small Language Model
Tianlin Li, Qian Liu, Tianyu Pang, Chao Du, Qing Guo, Yang Liu, Min Lin.

Your Large Language Model is Secretly a Fairness Proponent and You Should Prompt it Like One
Tianlin Li*, Xiaoyu Zhang*, Chao Du, Tianyu Pang, Qian Liu, Qing Guo, Chao Shen, Yang Liu.

ICSE 2024 (oral)

RUNNER: Responsible UNfair NEuron Repair for Enhancing Deep Neural Network Fairness
Tianlin Li*, Yue Cao*, Jian Zhang, Shiqian Zhao, Yihao Huang, Aishan Liu, Qing Guo, Yang Liu.
International Conference on Software Engineering

TOSEM 2023

Faire: Repairing Fairness of Neural Networks via Neuron Condition Synthesis
Tianlin Li, Xiaofei Xie, Jian Wang, Qing Guo, Aishan Liu, Lei Ma, Yang Liu.
ACM Transactions on Software Engineering and Methodology

ICML 2023

FAIRER: FAIRNESS AS DECISION RATIONALE ALIGNMENT
Tianlin Li, Qing Guo, Aishan Liu, Mengnan Du, Zhiming Li, Yang Liu.
International Conference on Machine Learning

Author Publications (*: co-first author, #:corresponding author) (continued)

IJCAI 2023 (oral)
Fairness via Group Contribution Matching
Tianlin Li, Zhiming Li, Anran Li, Mengnan Du, Aishan Liu, Qing Guo, Guozhu Meng, Yang Liu.
International Joint Conference on Artificial Intelligence

Inf. Sci. 2020
Understanding adversarial robustness via critical attacking route
Tianlin Li*, Aishan Liu*, Xianglong Liu, Yitao Xu, Chongzhi Zhang, Xiaofei Xie.
Information Sciences.

TOSEM 2021
NPC: Neuron Path Coverage via Characterizing Decision Logic of Deep Neural Networks
Xiaofei Xie*, **Tianlin Li***, Jian Wang, Lei Ma, Qing Guo, Felix Juefei-Xu, Yang Liu.
ACM Transactions on Software Engineering and Methodology

ICLR 2020
Knowledge consistency between neural networks
Ruofan Liang*, **Tianlin Li***, Longfei Li, Jing Wang, Quanshi Zhang.
International Conference on Learning Representations.

Preprint
A Mutation-Based Method for Multi-Modal Jailbreaking Attack Detection
Xiaoyu Zhang*, **Tianlin Li***, Cen Zhang, Yihao Huang, Xiaojun Jia, Xiaofei Xie, Yang Liu, Chao Shen.

LREC-Coling 2024
Unveiling Project-Specific Bias in Neural Code Models
Zhiming Li, Yanzhou Li, **Tianlin Li**#, Mengnan Du, Bozhi Wu, Yushi Cao, Xiaofei Xie, Yi Li, Yang Liu.
International Conference on Computational Linguistics, Language Resources and Evaluation

ICLR 2024
BadEdit: Backdooring Large Language Models by Model Editing
Yanzhou Li, **Tianlin Li**#, Kangjie Chen#, Jian Zhang, Shangqing Liu, Wenhan Wang, Tianwei Zhang, Yang Liu
International Conference on Learning Representations

Preprint
On the robustness of segment anything
Yihao Huang, Yue Cao, **Tianlin Li**#, Felix Juefei-Xu, Di Lin, Ivor W Tsang, Yang Liu, Qing Guo#.

ICLR 2024
IRAD: Implicit Representation-driven Image Resampling against Adversarial Attacks
Yue Cao, **Tianlin Li**, Xiaofeng Cao, Ivor Tsang, Yang Liu, Qing Guo
International Conference on Learning Representations

TMM 2024
Improving Deepfake Detection Generalization by Invariant Risk Minimization
Zixin Yin*, Jiakai Wang*, Yisong Xiao, Hanqing Zhao, **Tianlin Li**, Wenbo Zhou, Aishan Liu, and Xianglong Liu
IEEE Transactions on Multimedia

AAAI 2024 (oral)
FedMut: Generalized Federated Learning via Stochastic Mutation
Ming Hu, Yue Cao, Anran Li, Zhiming Li, Chengwei Liu, **Tianlin Li**, Mingsong Chen, Yang Liu.
AAAI Conference on Artificial Intelligence

AAAI 2024
Personalization as a Shortcut for Few-Shot Backdoor Attack against Text-to-Image Diffusion Models
Yihao Huang, Felix Juefei-Xu, Qing Guo, Jie Zhang, Yutong Wu, Hu Ming, **Tianlin Li**, Geguang Pu, Yang Liu.
AAAI Conference on Artificial Intelligence

Author Publications (*: co-first author, #:corresponding author) (continued)

- ASE 2023 — **Learning to Locate and Describe Vulnerabilities**
  Jian Zhang, Shangqing Liu, Xu Wang, **Tianlin Li**, Yang Liu.
  IEEE/ACM International Conference on Automated Software Engineering.
- ISSTA 2023 — **Latent Imitator: Generating Natural Individual Discriminatory**
  Yisong Xiao, Aishan Liu, **Tianlin Li**, Xianglong Liu.
  International Symposium on Software Testing and Analysis.
- TIP 2020 — **Interpreting and improving adversarial robustness of deep neural networks with neuron sensitivity**
  Chongzhi Zhang, Aishan Liu, Xianglong Liu, Yitao Xu, Hang Yu, Yuqing Ma, **Tianlin Li**.
  IEEE Transactions on Image Processing.

Research Interests

- Trustworthy AI — **Testing for AI-based software**;
  Fields of robustness, adversarial attack, OOD data, backdoor attack, and **fairness** on **AIGC models**.
- AI Interpretability — **Neuron-level interpretability**, especially the neuron representation of neural networks;
  Data-level interpretability, including instance-based interpretation for the learning process and distribution-aware interpretation.

Patents

- Aishan Liu, Xianglong Liu, **Tianlin Li**. 2020. **Methods and devices for determining critical attack path in neural networks.** C.N. Patent Application `202010888524.4`, filed August 2020.
- Aishan Liu, Xianglong Liu, **Tianlin Li**. 2020. **Neural network training methods and devices based on the critical path.** C.N. Patent Application `202010889881.2`, filed August 2020.
- Xiaofei Xie, **Tianlin Li**, Lei Ma, Yang Liu. 2022. **Network Model Testing Method Based on Key Decision Logic Design Test Coverage.** C.N. Patent Application `CN113255810B`, filed August 2022.

Miscellaneous Experience

Awards and Achievements

- 2024 — **DAAD AInet Fellowship**. Awarded by DAAD, German.
- 2022 — **3rd place in the AISG Trusted Media Challenge with 25K SGD cash prize, 2022**. Awarded by AI Singapore. News
- 2021-2024 — **AISG PhD Fellowship**. Awarded by AI Singapore.
- 2018 — **First Class Scholarship for Graduate Students**. Awarded by School of Computer Science and Engineering, Beihang University.
- 2017 — **First Class Scholarship for Graduate Students**. Awarded by School of Computer Science and Engineering, Beihang University.

Miscellaneous Experience (continued)

2016 **First Class Scholarship for Graduate Students**. Awarded by School of Computer Science and Engineering, Beihang University.

Certification

2018 **Outstanding Teaching Assistant**. Awarded by Beihang University.

Other Awards

2022 **NTU 3v3 Basketball Champion**.

2023 **Champion of the Four-Nation Embassy Basketball Tournament**.

Invited Talk/Interview

ICSE: Program Repair

Topic	Fairness via Group Contribution Matching
Date	4:00 pm - 4:15 pm on April 17, 2024
Location	Pequeno Auditório, Lisbon

Security of Large Language Models workshop

Topic	Model Security: Attacks and Defenses on Training Data and Model
Date	2 pm - 4 pm on September 18, 2023
Location	seminar room 1-1, academic building north (ABN), Nanyang Technological University

Audience	Tamasek;
	Infocomm Media Development Authority;
	Cyber Security Agency of Singapore;
	DSO National Laboratories;
	Certitude Singapore.

IJCAI: AI Ethics, Trust, Fairness

Topic	Fairness via Group Contribution Matching
Date	3:30 pm - 4:50 pm on August 23, 2023
Location	Almaty 6006, the Sheraton Grand Macao

Interview by CNA (Channel NewsAsia)

Topic	Insight 2023/2024 - How is TikTok changing Politics?
Date	2 pm - 4 pm on May 20, 2023
Location	Singapore Management University
News Link	CNA News

AISG Trusted Media Challenge Award Ceremony

Topic	Fake Bush Detector: a multi-modality deepfake detector
Date	4 pm - 5:30 pm on April 29, 2022
Location	innovation 4.0 (Seminar Room), 3 Research Link, Singapore 117602

Invited Talk/Interview (continued)

Audience

- Mr Tan Kiat How, Senior Minister of State, Ministry of Communications and Information & Ministry of National Development;
  Prof Ho Teck Hua, the fifth President of Nanyang Technological University;
  Mr Walter Fernandez, Editor-in-Chief and Chief Sustainability Officer, Mediacorp;
  Mr Eugene Leow, Head of Digital Media & Strategy, English/Malay/Tamil Media Group, Singapore Press Holdings;

Academic Services

Journals

- Reviewer of IEEE Transactions on Image Processing (TIP)
- Reviewer of Pattern Recognition (PR)
- Reviewer of Neurocomputing
- Reviewer of IEEE Internet of Things Journal (IoT)

Conferences

- Reviewers of AAAI, ICCV, CVPR, IJCAI.

Workshops

- Challenge Chair of The Art of Robustness: Devil and Angel in Adversarial Machine Learning at CVPR 2022.
- PC of CVPR, AAAI, MM workshops.