Yunzhe Tian

Phone: (+86) 130-2008-7266
Email: tianyunzhe@bjtu.edu.cn
Homepage: https://tianyunzhe.github.io

Summary

Yunzhe Tian is a Ph.D. candidate at the School of Cyberspace Science and Technology, Beijing Jiaotong University, advised by Prof. Wenjia Niu. From Nov, 2024 to Mar, 2025, he has been visiting Deakin University in Australia, working with Prof. Gang Li. He has contributed to over **10 high-quality** conference and journal papers, including publications in IJCAI, TETC, and TRE. His primary research area is **Trustworthy AI**, with a focus on developing **robust** and **explainable** learning models for broad AI applications, including computer vision, reinforcement learning, and wireless communication systems.

Education

Beijing Jiaotong University *Sep 2022 - Jul 2026 (expected)*
Beijing Key Laboratory of Security and Privacy in Intelligent Transportation
Ph.D. in Cyberspace Science and Technology Advisor: Prof. Wenjia Niu
◊ Dissertation: *Research on Interpretability-based Adversarial Example Detection and Restoration*

Beijing Jiaotong University *Sep 2020 - Jun 2022*
Beijing Key Laboratory of Security and Privacy in Intelligent Transportation
Master in Artificial Intelligence Advisor: Prof. Wenjia Niu
◊ Dissertation: *Research on Robustness of Graph Neural Network Based on Curriculum Learning*
(Excellent Master Dissertation of Beijing Jiaotong University)

Beijing Information Science & Technology University *Sep 2016 - Jun 2020*
Bachelor in Information System & Information Management (*GPA: 4.02*)
◊ Dissertation: *Research on News Entity Open Relation Extraction* Advisor: Dr. Peng Zhang
(Work done during visiting at KEG@THU; Excellent Undergraduate Dissertation of Beijing City)

Visiting Experience

Deakin University *Nov 2024 - Mar 2025*
Team for Universal Learning and Intellgient Processing (TULIP)
Visiting Ph.D. Advisor: Prof. Gang Li

Tsinghua Univerisity *Sep 2018 - Jun 2020*
Knowledge Engineering Group
Visiting Student Advisors: Prof. Juanzi Li and Dr. Peng Zhang

Research Interests

My research centers on trustworthy AI for real-world applications, with an emphasis on robustness, and explainability, and their interconnections.

- **AI robustness.** I have worked on exploring adversarial attacks (both evasion and poisoning) across multiple learning paradigms and models, including continual learning, reinforcement learning, graph neural networks, and spiking neural networks. I have developed and will continue to explore robust learning systems based on model explanations, curriculum learning, and causal learning, with applications in safe wireless communication systems, and trustworthy reinforcement learning systems.
- **AI explainability.** I have explored explainability for real-world AI, including proposing model-agnostic methods and quantitative explanation evaluation frameworks for wireless communication systems. My long-term goal is to investigate the connection between explainability and robustness, using explanations to improve failure diagnosis and guide adversarial defenses (including detection, restoration, and retraining) that strengthen the reliability and trustworthiness of deployment-critical AI systems.

Selected Publications

First Author

[TETC'25] **Yunzhe Tian**, Yike Li, Kang Chen, Zhenguo Zhang, Endong Tong, Jiqiang Liu, Fangyun Qin, Zheng Zheng, and Wenjia Niu. Towards Label-Efficient Deep Learning-based Aging-related Bug Prediction with Spiking Convolutional Neural Networks. In *Transactions on Emerging Topics in Computing, 2025.* **(IF=5.4)**

[TRE'24] **Yunzhe Tian**, Dongyue Xu, Endong Tong, Rui Sun, Kang Chen, Yike Li, Thar Baker, Wenjia Niu, and Jiqiang Liu. Toward Learning Model-Agnostic Explanations for Deep Learning-Based Signal Modulation Classifiers. In *IEEE Transactions on Reliability, 2024.* **(IF=5.7)**

[IJCAI'23] Yike Li, **Yunzhe Tian (co-first author)**, Endong Tong, Wenjia Niu, and Jiqiang Liu. Robust Reinforcement Learning via Progressive Task Sequence. In *Proceedings of the 32nd International Joint Conference on Artificial Intelligence (IJCAI 2023), 2023.* **(CCF-A)**

[ISSRE'23] **Yunzhe Tian**, Yike Li, Kang Chen, Endong Tong, Wenjia Niu, Jiqiang Liu, Fangyun Qin, Zheng Zheng. Mitigating Overfitting for Deep Learning-based Aging-related Bug Prediction via Brain-inspired Regularization in Spiking Neural Networks. In *IEEE 34th International Symposium on Software Reliability Engineering Workshops (ISSREW 2023), 2023.* **(CCF-B workshop)**

[AutoSec'21] **Yunzhe Tian**, Yike Li, Yingxiao Xiang, Wenjia Niu, Endong Tong, and Jiqiang Liu. Curricular Reinforcement Learning for Robust Policy in Unmanned CarRacing Game. In *NDSS 2021, Workshop on Automotive and Autonomous Vehicle Security (AutoSec).* **(CCF-A workshop)**

[SCN'21] **Yunzhe Tian**, Jiqiang Liu, Endong Tong, Wenjia Niu, Liang Chang, Qi Alfred Chen, Gang Li, and Wei Wang. Towards Revealing Parallel Adversarial Attack on Politician Socialnet of Graph Structure. In *Security and Communication Networks (SCN), 2021.* **(CCF-C)**

[ICPADS'21] **Yunzhe Tian**, Yingdi Wang, Endong Tong, Wenjia Niu, Liang Chang, Qi Alfred Chen, Gang Li, and Jiqiang Liu. Exploring Data Correlation between Feature Pairs for Generating Constraint-based Adversarial Examples. In *The IEEE 26th International Conference on Parallel and Distributed Systems (ICPADS 2020), 2020.* **(CCF-C)**

Co-Author

[计算机研发'25] 徐冬月, **田蕴哲**, 陈康, 李轶珂, 吴亚伦, 童恩栋, 牛温佳, 刘吉强, 史忠植. 面向信号调制识别的对抗攻击与防御综述. *计算机研究与发展, 2025.*

[DSC'25] Jingqi Jia, **Yunzhe Tian**, Xinyi Zhao, Xiangyu Shi, Ping Ye, Xiaoshu Cui, Yanfeng Gu, Xingyu Wu, Dianjing Cheng, and Wenjia Niu. Graph Neural Network-Enhanced Semantic Invariant Robust Watermarking for Large Language Models. In *The IEEE 9th International Conference on Data Science in Cyberspace (DSC 2025), 2025.* **(Outstanding Paper)**

[APL'25] Xingyu Wu, **Yunzhe Tian**, Yuanwan Chen, Ping Ye, Xiaoshu Cui, Jingqi Jia, Shouyang Li, Jiqiang Liu, and Wenjia Niu. CurriculumPT: LLM-Based Multi-Agent Autonomous Penetration Testing with Curriculum-Guided Task Scheduling. In *Applied Sciences, 2025.*

[ASE'25] Jiayin Song, Yike Li, **Yunzhe Tian**, Haoxuan Ma, Honglei Li, Jie Zuo, Jiqiang Liu, and Wenjia Niu. Investigating the bugs in reinforcement learning programs: Insights from Stack Overflow and GitHub. In *Automated Software Engineering, 2025.*

[KSEM'24] Jiayin Song, Yike Li, **Yunzhe Tian**, Xingyu Wu, Qiong Li, Endong Tong, Wenjia Niu, Zhenguo Zhang, and Jiqiang Li. Knowledge-Driven Backdoor Removal in Deep Neural Networks via Reinforcement Learning. In *The 17th International Conference on Knowledge Science, Engineering and Management (KSEM 2024), 2024.*

[TGCN'22] Yike Li, Wenjia Niu, **Yunzhe Tian**, Tong Chen, Zhiqiang Xie, Yalun Wu, Yingxiao Xiang, Endong Tong, Thar Baker, and Jiqiang Liu. Multiagent Reinforcement Learning-Based Signal Planning for Resisting Congestion Attack in Green Transportation. In *IEEE Transactions on Green Communications and Networking (TGCN), 2022.*

[TGCN'21] Endong Tong, Wenjia Niu, **Yunzhe Tian**, Jiqiang Liu, Thar Baker, Sandeep Verma, and Zheli Liu. A Hierarchical Energy-efficient Service Selection Approach with Qos Constraints for Internet of Things. In *IEEE Transactions on Green Communications and Networking (TGCN), 2021.*

[JCSC'21] Yingdi Wang, **Yunzhe Tian**, Jiqiang Liu, Wenjia Niu, and Endong Tong. A Training-Based Identification Approach to VIN Adversarial Examples in Path Planning. In *Journal of Circuits, Systems and Computers, 2021.*

[TST'21] Yike Li, **Yunzhe Tian**, Endong Tong, Wenjia Niu, Yingxiao Xiang, Tong Chen, Yalun Wu, and Jiqiang Liu.Curricular Robust Reinforcement Learning via GAN-based Perturbation through Continuously-scheduled Task Sequence. In *TSINGHUA Science and Technology (TST), 2021.*

[计算机学报'21] 王硕汝, 牛温佳, 童恩栋, 陈彤, 李赫, **田蕴哲**, 刘吉强, 韩臻. 强化学习离线策略评估研究综述. *计算机学报, 2021.*

[SCN'21] Xinyu Huang, **Yunzhe Tian**, Yifei He, Endong Tong, Wenjia Niu, Chenyang Li, Jiqiang Liu, and Liang Chang. Exposing Spoofing Attack on Flocking-based Unmanned Aerial Vehicle Cluster: A Threat to Swarm Intelligence. In *Security and Communication Networks (SCN), 2020.*

[INSCRYPT'20] Bowei Jia, **Yunzhe Tian**, Di Zhao, Xiaojin Wang, Chenyang Li, Wenjia Niu, Endong Tong, and Jiqiang Liu. Bidirectional Rnn-based Few-shot Training for Detecting Multi-stage Attack. In *The 16th International Conference on Information Security and Cryptology (INSCRYPT 2020), 2020.*

[CCKS'20] Qinghua Wen, **Yunzhe Tian**, Xiaohui Zhang, Ruoyun Hu, Jinsong Wang, Lei Hou, and Juanzi Li. Type-aware Open Information Extraction via Graph Augmentation Model. In *China Conference on Knowledge Graph and Semantic Computing (CCKS 2020), 2020.*

Project Experience

Project PI

Research on Interpretability of Signal Recognition Based on Residual Attention Networks
15,000 RMB, *Apr. 2023 - Mar. 2025*

The Fundamental Research Funds for the Central Universities of China (Grant No. 2023YJS031). (Awarded **Excellent Completion**).

- Proposed a novel model-agnostic explainer for the predictions of black-box signal classifier.
- Developed the first generic quantitative explanation evaluation framework for signal classification.
- **Research Outcome**: A journal paper published in *IEEE Transactions on Reliability, 2024.*

Project Member

Testing of Deep Reinforcement Learning Software Systems: An Approach Driven by Coverage of Markov Decision Sequence Relation 500,000 RMB, *Jan. 2024 - Present*

The National Natural Science Foundation of China (Grant No. 62372021).

- Constructed a dataset of real-world RL-related bugs from *Stack Overflow* and *Github.*
- Proposed label-efficient aging-related bug prediction with spiking convolutional neural networks.
- **Research Outcome**: A paper published in *IEEE Trans on Emerging Topics in Computing, 2025.*

Security Threat Analysis of Low-Power and Cold-Start Wireless Communication Systems
200,000 RMB, *Apr. 2023 - Present*

The Fundamental Research Funds for the Central Universities of China (Grant No. 2023JBZY036).

- Proposed an explainability-based framework for detection and restoration of adversarial IQ signals.
- Optimized spiking neural networks for low-power edge signal recognition on neuromorphic chips.
- **Research Outcome**: A paper submitted to *IEEE Internet of Things Journal, 2025.*

Research on Multi-Agent Collaborative Defense Against Data Poisoning Attacks in Intelligent Traffic Signal Systems 600,000 RMB, *Oct. 2019 - Dec. 2023*

The National Natural Science Foundation of China (Grant No. 61972025).

- Deployed the I-SIG system in a real-world setting on Xinshi Trial Road in Shijiazhuang, China.
- Developed a robust reinforcement learning algorithm based on curriculum task generation.
- **Research Outcome**: A conference paper published in *IJCAI 2023.*

Teaching & Mentoring Experience

Teaching Assistant *Feb. 2023 - Jun. 2023*
M602031B: Situation Awareness of Cyberspace Security. Instructor: Prof. Wenjia Niu

Teaching Assistant *Jun. 2023 - Jul. 2023*
80S504Q: Information Security Professional Practice and Training. Instructor: Prof. Wenjia Niu

Teaching Assistant *Feb. 2024 - Jun. 2024*
M402055B: Artificial Intelligence Security. Instructor: Prof. Wenjia Niu

Guest Lecturer *Mar. 2025*
Towards Revealing Persistent Universal Attacks in Continual Learning Instructor: Prof. Gang Li
Deakin Cyber 2025 Seminar Series.

Guest Lecturer *Jun. 2025*
Is AI Safe? Adversarial Attack and Defense Instructor: Prof. Wenjia Niu
Seminar on Cybersecurity under Global Security Intiative.

Research Advising and Mentoring

Team leader of the TrustAI group, a subgroup within the THETA Lab led by Prof. Wenjia Niu.
I am fortunate to mentor brilliant teammates:

- Shiyao Chen (BJTU M.S.) *Sep. 2021 - Jul. 2024*
  Thesis: Research of Robust Reinforcement Learning Based on Anomalous State Enhancement
  Awarded **Outstanding Master Thesis of Beijing Jiaotong University**

- Zhenglong Liu (BJTU B.S., now M.S. at USC) *May. 2022 - Jul. 2024*
  Project: Unity3D-Based Knowledge Competition System
  Awarded **National-level College Student Innovative Training Program**
- Xinyi Zhao (BJTU B.S., now Ph.D. at BJTU) *Sep. 2021 - Jul. 2025*
  Thesis: Research on Adversarial Attacks and Evaluation for SNN-Based Modulation Recognition
- Dongyue Xu (BJTU B.S.) *Sep. 2022 - Jul. 2025*
  Thesis: Research on Security Defense Methods for Self-Supervised Models in Signal Recognition
- Kang Chen (BJTU B.S.) *Sep. 2022 - Jul. 2024*
  Thesis: Ensemble Learning and Model Security of Spiking Neural Networks for Signal Recognition
- Rui Sun (BJTU B.S.) *Sep. 2021 - Jul. 2023*
  Thesis: Research on Signal Modulation Recognition Based on Residual Attention Network

Academic Experience

Program Committee (PC) member of **AAAI 2026**, **DSN 2025**

Journal Reviewer of **Journal of Information and Knowledge Management (JIKM)**

Journal Reviewer of **Journal of Supercomputing**

Journal Reviewer of **Internet of Things (IoT)**

Oral Presentation in **AUTODRIVING TECH TALK @ BCTF 2022**

Oral Presentation in **AutoSec Workshop @ NDSS'21**

Oral Presentation in **Inscrypt 2020**, Guangzhou, China

Oral Presentation in **ICPADS 2020**, Hong Kong, China

Selected Honors and Awards

First-class Ph.D. Scholarship of Beijing Jiaotong University (Top 10%).	*2022, 2023, 2024, 2025*
Second Prize in the 34th Huiguang Cup Academic Cultural Festival (Academic Poster Track)	*2024*
Fourth Place in IEEE Trojan Removal Competition (IEEE TRC @ ICLR 2023).	*2023*
Excellent Team in DataCon Big Data Security Analysis Competition.	*2023*
First Prize in Vulnerability Mining Contest for Olympic Winter Games Beijing.	*2022*
Second Prize in DEF CON 30 Contest AutoDriving CTF.	*2022*
Excellent Master Thesis of Beijing Jiaotong University (Top 2%).	*2022*
Excellent Graduate of Beijing Jiaotong University (Top 5%).	*2022*
Second Prize in 第二届全国分布式靶场安全技能大赛	*2021*
Second Prize in DEF CON 29 Contest AutoDriving CTF.	*2021*
Excellent Undergraduate of Beijing City (Top 1%).	*2020*
Excellent Undergraduate Thesis of Beijing City (Top 1%).	*2020*
Excellent Undergraduate of Beijing Information Science & Technology University (Top 5%).	*2020*
Excellent Undergraduate Thesis of Beijing Information Science & Technology University.	*2020*
National Scholarship, The Ministry of Education of China (Top 1%).	*2019*
President Scholarship of Beijing Information Science & Technology University (Top 1%).	*2019*

Skills

Programming Languages:	Proficient with Python, Java, Shell, SQL
Deep Learning Software Stacks:	Pytorch, Tensorflow, Scikit-learn, NumPy, Pandas
Technical & Collaborative Tools:	Git, Latex, Huggingface, Visio, Zotero
Languages:	Mandarin Chinese (Native), English (CET-6: 524, IELTS: 6.5)

Academic References

- Prof. Wenjia Niu (niuwj@bjtu.edu.cn), Beijing Jiaotong University
- A/Prof. Endong Tong (edtong@bjtu.edu.cn), Beijing Jiaotong University