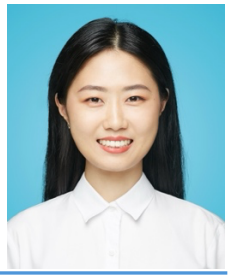

# 李轶珂

中共党员 | 13718232088 | yikeli@bjtu.edu.cn | https://li-yike.github.io

## 教育背景

- **北京交通大学** **网络空间安全 | 硕士，博士** **2019.09 – 至今**
  智能交通数据安全与隐私保护技术北京市重点实验室 导师：牛温佳教授
  - 荣誉奖项：2022, 2023, 2024 年博士研究生一等奖学金、2021 年博士新生奖学金。
- **合肥工业大学** **信息安全 | 学士** **2014.09 – 2018.06**
  - 荣誉奖项：校级专项奖学金、校级学业奖学金、校“互联网+”创新创业竞赛优胜奖。

## 学术成果

- **研究方向：**
  - 人工智能安全，包括：模型鲁棒性，模型泛化性，模型隐私安全。
  - 应用场景：强化学习系统隐私攻击与防护、智能交通系统中的数据投毒对抗攻击与防护。
- **论文发表：**
  - **第一作者 5 篇（包括 CCF-A 人工智能顶级会议 1 篇、CCF-A 信息安全顶级期刊 1 篇）如下。**
  - **第二作者 4 篇，第三作者 5 篇（包括中科院一区期刊 1 篇、CCF-B 会议 1 篇）。**

**Yike Li,** Yunzhe Tian, Endong Tong, Wenjia Niu and Jiqiang Liu. Robust Reinforcement Learning via Progressive Task Sequence. In International Joint Conference on Artificial Intelligence (IJCAI), 2023. **(CCF-A，录取率 14.1% = 643/4566)**

**Yike Li,** Jiayin Song, Yunzhe Tian, Endong Tong, Yuling Liu, Guozhu Meng, Yalun Wu, Jianhua Li, Wenjia Niu, and Jiqiang Liu. Towards Preventing Imitation Learning Attack via Policy Confusion Defense. (Under Second Review) Summited to IEEE Transactions on Dependable and Secure Computing (TDSC), 2023. **(CCF-A, SCI 检索，中科院一区期刊, IF=7.0) (二审中)**

**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), 2022. **(SCI 检索，中科院一区期刊, IF=5.2)**

**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.**( SCI 检索，中科院二区期刊, IF=5.3)**

**Yike Li,** Yingxiao Xiang, Endong Tong, Wenjia Niu, Bowei Jia, Long Li, Jiqiang Liu, and Zhen Han. An Empirical Study on GAN-Based Traffic Congestion Attack Analysis: A Visualized Method. In Wireless Communications and Mobile Computing (WCMC), 2020. **(CCF-C, SCI 检索，中科院四区期刊)**

- **学术活动及获奖：**
  - 交流经历：于 2024.02-2024.08 前往澳大利亚迪肯大学公派交流，校外导师：Prof. Gang Li，Team for Universal Learning and Intelligent Processing (TULIP).
  - 担任学术期刊审稿人：Journal of Information and Knowledge Management (JIKM)
    Journal of Intelligent & Fuzzy Systems (JIFS)
  - 学术海报：获第 34 届北京交通大学慧光杯学术文化节学术海报组二等奖
  - 学术报告：CTCIS 2021 会议（保定，中国）口头报告
    IJCAI 2023 会议（澳门，中国）口头报告
  - 课程助教：信息安全综合训练 I (80S504Q) 2021.06 – 2021.07
    网络空间安全态势感知 I (M602031B) 2021.02 – 2022.04
    人工智能安全 (M402055B) 2023.02 – 2023.08

## 项目经历

**■ (主持) 基于课程式任务生成的强化学习鲁棒性研究 (优秀结题)** **2022.04 – 2024.03**

- 项目来源：北京交通大学研究生创新项目。
- **参与情况：独立完成。**
  (1) 首次提出“最大-期望”的鲁棒学习范式，有效解决传统“最大-最小”范式中策略过度保守问题。
  (2) 提出基于渐进式多任务学习的鲁棒训练算法，来求解所提出的“最大-期望”问题。
  (3) 多个环境中的测试结果显示，所提议方法框架在最坏扰动下使策略奖励平均提升 23.21%。
- **项目成果：**第一作者发表论文 1 篇（IJCAI 2023)。

**■ (参与) 基于智能交通信号灯系统中数据投毒攻击的多主体协同防护研究** **2019.10 – 2023.12**

- 项目来源：国家自然科学基金“面上”项目。
- **参与情况：负责文档设计、参与工程实现。**
  (1) 平台搭建(负责)： 基于美国交通部开源项目 MMITSS，在本地设计并实现系统框架。
  (2) 攻击威胁分析(负责)：实施针对目标路口的数据投毒攻击，并量化在固定时间内的拥堵程度。
  (3) 防护框架设计(参与)：基于多智体强化学习算法构建多路口信号规划算法，该协同规划机制显著提升了系统对拥塞攻击的鲁棒性。测试结果显示，攻击场景下拥堵程度降低 64.33%。
- 项目成果: (1)在石家庄新华中路部署测试，利用雷视一体机和东土正创信号机实现最优配时规划。
  (2)第一作者发表论文 2 篇 (TGCN 2022, WCMC 2020)，第二作者 1 篇 (软件学报 2021)。

**■ (参与) 马氏决策序列关系覆盖驱动的深度强化学习软件测试研究** **2024.01 – 至今**

- 项目来源：国家自然科学基金“面上”项目（合作）。
- **参与情况：负责文档设计、参与工程实现。**
  (1) 构建数据集(参与)：抓取强化学习代码缺陷数据，进行大语言模型特征初筛、人工二次筛选。
  (2) 测试方法设计(负责)：提出面向强化学习系统的蜕变测试方法，实现高效测试用例生成工具。
- 项目成果：(1) 构建首个强化学习系统缺陷数据集，包含 4000 条高质量标记样本。
  (2) 第二作者发表论文 2 篇 (KSEM 2024, 1 篇在审)。

**■ (参与) 大规模 XXX 安全知识图谱** **2022.05 – 2022.12**

- 项目来源：中国人民解放军联参某部队项目。
- **参与情况：负责文档设计、前期系统设计。**
  负责整体需求分析，并设计系统功能模块、设计图谱 Knowledge Schema。
- 项目成果: (1)图谱系统成功部署，获表扬信和系统应用证明（4/10）; (2)系统框架应用于二期项目。

## ■ 学科竞赛

- DataCon 2023 大数据安全分析竞赛 AI 安全赛道｜卓越团队奖（第四名） 2023
- 2022 北京冬奥会系统安全测试大赛｜团队一等奖 2022
- 第 17 届中国研究生数学建模竞赛（华为杯）｜二等奖 2020

## 个人技能

- **英语水平：**CET-4 564 分，CET-6 508 分，具备良好的英文听说读写能力。
- **专业技能：**(1) 熟练使用深度学习框架：PyTorch, TensorFlow, RLlib, Tianshou, Stable-Baselines,
  (2) 熟练使用编程语言：Python, C++,
  (3) 熟练使用专业工具：Git, Latex, Visio, Zotero 等,
  (4) 熟悉人工智能方法及原理、攻击方法及原理(后门攻击、隐私攻击、数据欺骗攻击等)、防御方法及原理（混淆防御、拟态防御等)。
- **项目协调能力：**熟悉项目进展周期，善于发现问题并解决问题。
  (1) 快速调研当前领域现状，包括：学术研究进展、业界主流应用、常用开源工具等,
  (2) 快速复现现有工作，包括：开源代码、开源工具等,
  (3) 围绕需求设计方法框架和评估指标，并进行实验分析。