ZHI YANG

+86-138-2874-1098 | 12332454@mail.sustech.edu.cn | ruayz.github.io
Research area: Trustworthy Machine Learning.

Education

- **Southern University of Science and Technology** *Sep. 2023 - Present*
  Master's degree in Computer Science and Technology (expected)
  - Supervisor: Changwu Huang, Ke Tang, Xin Yao
  - GPA: 3.6/4.0
  - Core courses: Advanced Artificial Intelligence (B), Advanced Algorithms (A-), Numerical Methods(B), Human Brain Intelligence and Machine Intelligence (A-), etc.
- **Yunnan University** *Sep. 2019 - Jun. 2023*
  Bachelor's degree in Software Engineering
  - GPA: 3.8/4.0 (Top 2%)
  - Core courses: Advanced Mathematics (92), Discrete Mathematics (94), Probability and Statistics (95), Computer Networks (95), Software Engineering (93), etc.

Publications

C=Conference, S=In Submission

[S.1] Zhi Yang, Changwu Huang, Ke Tang, Xin Yao. **On the Fairness of Privacy Protection: Measuring and Mitigating the Disparity of Group Privacy Risk for Differentially Private Machine Learning**. Manuscript under review, 2025.

[C.1] Zhi Yang, Changwu Huang, Xin Yao. **Towards Private and Fair Machine Learning: Group-Specific Differentially Private Stochastic Gradient Descent with Threshold Optimization**. *The International Conference on Neural Information Processing (ICONIP)*. Singapore: Springer Nature Singapore, 2024: 66-80.

[C.2] Zhi Yang, Ziming Wang, Changwu Huang, Xin Yao. **An Explainable Feature Selection Approach for Fair Machine Learning**. *International Conference on Artificial Neural Networks (ICANN)*. Cham: Springer Nature Switzerland, 2023: 75-86.

Research Experience

- **On the Fairness of Privacy Protection [S.1]**
  MASTER'S THESIS []
  - Summary: While fairness in conventional machine learning and differentially private machine learning (DPML) has been extensively studied, the fairness of privacy protection across groups remains underexplored. Existing methods for assessing group privacy risks are either insufficiently accurate or computationally expensive. To address these limitations, we propose a novel membership inference game that efficiently approximates worst-case privacy risks, enabling stricter and more reliable assessment of the disparity in group privacy risks. To further enhance fairness, we introduce a group-specific adaptive gradient clipping strategy for DP-SGD, which effectively reduces disparities in group privacy risks.
- **Towards Private and Fair Machine Learning [C.1]**
  MASTER'S THESIS []
  - Summary: Recent research has highlighted the challenges of integrating differential privacy (DP) with group fairness. One line of work focuses on mitigating accuracy disparities across sensitive groups introduced by DP mechanisms, while another seeks to preserve outcome fairness in DP-trained models. However, these two objectives often conflict, and existing methods typically address them in isolation. To tackle both problems simultaneously, we propose a group-specific DP-SGD training framework combined with classification threshold optimization. Our approach jointly reduces accuracy disparity and achieves outcome fairness.
- **An Explainable Feature Selection Approach [C.2]**
  BACHELOR'S THESIS
  - Summary: In this work, we propose an explainable feature selection (ExFS) method to improve the fairness of ML by recursively eliminating features that contribute to unfairness based on the feature attribution explanations (FAE) of the model's predictions.

Honors and Awards

• Excellent teaching assistant	*Sep. 2024*
• Top research assistance scholarship	*Sep. 2024*
• National Scholarship	*Dec. 2022*
• The 8th Internet plus Innovation and Entrepreneurship Competition, National Silver Award	*Nov. 2022*
• University-level first-class scholarship	*Dec. 2021*
• University-level first-class scholarship	*Dec. 2020*