Yanyun Wang 王彦云

DOB: 15/05/2000
Tel: +86 18678857129
Email: ywang856@connect.hkust-gz.edu.cn
Homepage: https://yywang.netlify.app

Education

M.Phil. Artificial Intelligence, The Hong Kong University of Science and Technology (Guangzhou) **09/2024 – Now**

- GPA: 4.00/4.3 **- Supervisor:** Prof. Li Liu **- Co-supervisor:** Prof. Yi R. (May) Fung (HKUST)

M.Sc. Computer Science, The University of Hong Kong **09/2022 – 10/2023**

- GPA: 3.35/4.3

- Degree Project: *Towards Robust Speaker Recognition through Crafting Imperceptible Adversarial Speech Samples* (Grade: A+)

B.Eng. Software Engineering, East China Normal University (985 / Double First Class) **09/2018 - 06/2022**

- GPA: 3.52/4.0 **- Ranking:** 30/186 (16.13%)

- Degree Thesis: *RNN Adversarial Samples Generation Approach based on Weighted Finite Automaton Abstraction* (Grade: A)

Publications & Preprints

*: Corresponding author(s)

As The 1st Author:

[Preprint] ***Robust Alignment: Harmonizing Accuracy and Robustness in Adversarial Training***
Yanyun Wang, Qingqing Ye, Li Liu, Zi Liang, Haibo Hu*
Under review (TIFS'25)

[Preprint] ***New Paradigm of Adversarial Training: Releasing Accuracy-Robustness Trade-Off via Dummy Class***
Yanyun Wang, Li Liu*, Zi Liang, Yi R. (May) Fung, Qingqing Ye, Haibo Hu
Under review (ICLR'25 6665 rej -> NeurIPS'25)

[ICCV'25] ***Failure Cases Are Better Learned But Boundary Says Sorry: Facilitating Smooth Perception Change for Accuracy-Robustness Trade-Off in Adversarial Training***
Yanyun Wang, Li Liu*
International Conference on Computer Vision (CCF-A) **Accepted to appear**

[ECAI'24] ***TSFool: Crafting Highly-Imperceptible Adversarial Time Series through Multi-Objective Attack***
Yanyun Wang, Dehui Du*, Haibo Hu*, Zi Liang, Yuanhao Liu
European Conference on Artificial Intelligence (CCF-B) **Oral**

[SMC'23] ***Meta Pattern Concern Score: A Novel Evaluation Measure with Human Values for Multi-classifiers***
Yanyun Wang, Dehui Du*, Yuanhao Liu
IEEE International Conference on Systems, Man, and Cybernetics (CCF-C)

As Co-Author:

[Preprint] ***Virus Infection Attack on LLMs: Your Poisoning Can Spread "VIA" Synthetic Data***
Zi Liang, Qingqing Ye, Xuan Liu, **Yanyun Wang**, Jianliang Xu, Haibo Hu*
Under review (NeurIPS'25)

[Preprint] ***BackdoorDM: A Comprehensive Benchmark for Backdoor Learning on Diffusion Model***
Weilin Lin, Nanjun Zhou, **Yanyun Wang**, Jianze Li, Hui Xiong, Li Liu*
Under review (NeurIPS'25)

[ACL'25] ***"Yes, My LoRD." Guiding Language Model Extraction with Locality Reinforced Distillation***

Zi Liang, Qingqing Ye, **Yanyun Wang**, Sen Zhang, Yaxin Xiao, RongHua Li, Jianliang Xu, Haibo Hu*

Annual Meeting of the Association for Computational Linguistics (CCF-A) **Main conference**

[AAAI'22] ***Efficient Adversarial Sequence Generation for RNN with Symbolic Weighted Finite Automata***

Mingjun Ma, Dehui Du*, Yuanhao Liu, **Yanyun Wang**, Yiyang Li

SafeAI Workshop @ AAAI Conference on Artificial Intelligence (CCF-A) **Best paper award nomination**

Research & Intern Experience

Research Assistant, The Hong Kong Polytechnic University **10/2023 - 06/2024**

- **Laboratory:** Applied Security, Trust And Privacy Lab for Enterprise (ASTAPLE) - **Supervisor:** Prof. Haibo Hu

Research Assistant (part-time), East China Normal University **12/2021 - 09/2023**

- **Laboratory:** Shanghai Key Laboratory of Trustworthy Computing - **Supervisor:** Prof. Dehui Du

Algorithm Engineer, Ping An Technology Co., Ltd **08/2021 - 11/2021**

- **Department:** NLP Innovation Research and Development Department, OLATOP Knowledge Graph Team

Software Development Engineer, Dareway Software Co., Ltd **07/2020 - 08/2020**

Extracurricular Practice

Summer Workshop Student, School of Computing, National University of Singapore **05/2021 - 07/2021**

- **Topic:** AI/ML for Financial Services - **Grade:** A+
- **Project:** *Portfolio Management - Based on LSTM Models and Optimal Combination*

Participant, Shanghai College Students' Innovation and Entrepreneurship Training Program **10/2019 - 10/2020**

- **Project:** *Jiangnan Mengxun* – a 2.5D indie word adventure game - **Grade:** Provincial Level - B

Associate Director, Human Resource Center of Students' Union **06/2019 - 06/2020**

Honors & Awards

Excellent Graduate Award, East China Normal University **2022**

Excellent Bachelor's Degree Thesis Award, Software Engineering Institute, East China Normal University **2022**

People's Choice Award, School of Computing Summer Workshop, National University of Singapore **2021**

Excellent Undergraduate Student, East China Normal University **2021**

Second-class Scholarship, East China Normal University **2021**

Excellent Undergraduate Student, East China Normal University **2020**

First-class Scholarship, East China Normal University **2020**

Services

Conference Reviewer: CVPR (2025, 2024), ICCV (2025), AAAI (2025)

Journal Reviewer: TKDE