Yifan Lan

https://yifan-lan.github.io/ · Google Scholar · yifanlan@psu.edu · (+1) 814-280-0000

Education

The Pennsylvania State University Sep 2025 – present

PhD Student

Huazhong University of Science and Technology (HUST), Wuhan, China Sep 2021 – Jun 2025

Bachelor of Engineering in Artificial Intelligence

- GPA: 4.29/5.0
- Awards: Freshman Scholarship (2/30, 2022), "Self-Strengthening" Scholarship (1/27, 2023)

National University of Singapore, Singapore May 2023 – Jul 2023

NUS SOC Summer Workshop

Publications

- **Yifan Lan**, Hanyu Wang, Lu Lin, Jinghui Chen. "Skill Transfer: Adapting Agent Skills from Strong to Weak Agents." Accepted to ***NeurIPS 2026 TTCL Workshop*** (Poster), 2026. `https://openreview.net/forum?id=ag6fz5QxZ7`
- **Yifan Lan**, Yuanpu Cao, Weitong Zhang, Lu Lin, Jinghui Chen. "Phi: Preference Hijacking in Multi-modal Large Language Models at Inference Time." Accepted to ***EMNLP 2025 Oral***, 2025. `https://openreview.net/forum?id=e0uaB95224`
- **Yifan Lan**, Cai Xin, Jun Cheng, Shan Tan. "Mixture of Balanced Information Bottlenecks for Long-Tailed Visual Recognition." Accepted to ***TMLR***, 2025. `https://openreview.net/forum?id=9eiALSuZGA`

Preprints

- **Yifan Lan**, Zhongye Liu, Yurui Chang, Hanyu Wang, Yongkang Du, Lu Lin, Jinghui Chen. "DecisionRobust: Benchmarking Robustness for JEV Style Decision Models Across Text and Vision." *arXiv preprint arXiv:2610.19118*, 2026. `https://arxiv.org/abs/2610.19118`
- Hanyu Wang, **Yifan Lan**, Bochuan Cao, Lu Lin, Jinghui Chen. "SkillGrad: Optimizing Agent Skills Like Gradient Descent." *arXiv preprint arXiv:2605.27760*, 2026. `https://arxiv.org/abs/2605.27760`
- **Yifan Lan**, Yuanpu Cao, Hanyu Wang, Lu Lin, Jinghui Chen. "The Illusion of Reasoning: Exposing Evasive Data Contamination in LLMs via Zero-CoT Truncation." *arXiv preprint arXiv:2605.21856*, 2026. `https://arxiv.org/abs/2605.21856`

Honors & Awards

- **Fund for Excellence in Graduate Recruitment (FEGR) Award**, The Pennsylvania State University, 2025.
- **Self-Strengthening Scholarship**, Huazhong University of Science and Technology, 2023.
- **Freshman Scholarship**, Huazhong University of Science and Technology, 2022.

Research Support

- **NSF ACCESS Computational Resource Allocation**, 2026.

Research Interest

AI Agent, Reasoning, Trustworthy Machine Learning

Research Experiences

Research Assistant, College of IST, The Pennsylvania State University Jul 2024 – Present

Trustworthiness of Code LLM **Advisor: Professor Jinghui Chen**

- Investigated the risk and vulnerability of Code LLMs.
- Proposed a method using steering vectors to induce Code LLM to generate vulnerable code.

Preference Hijacking of MLLM **Advisor: Professor Jinghui Chen**

- Investigated the trustworthiness and persona of multimodal large language models (MLLMs).
- Uncover a new safety risk of MLLMs: the output preference of MLLMs can be arbitrarily manipulated by carefully optimized images.
- Proposed a method for manipulating the MLLM response preferences using a preference hijacked image. Our method works at inference time and requires no model modifications.
- Introduced a universal hijacking perturbation – a transferable component that can be embedded into different images to hijack MLLM responses toward any attacker-specified preferences

Research Assistant, College of CS, Northeastern University, Boston Mar 2024 – Aug 2024

Enhancing Long-Tailed Recognition with Retrieval Augmentation **Advisor: Professor Huaizu Jiang**

- Developed a binary labeling method to optimize weight assignment in retrieved data, significantly enhancing accuracy and effectiveness.
- Proposed and implemented three information merging strategies: label merging, logit merging, and feature merging. Analyzed experimental results, identifying distinct characteristics of each approach.
- The three information merging methods, enhanced by a binary labeling weight assignment module, outperformed previous retrieval augmentation techniques without requiring ViT fine-tuning.

Research Assistant, School of AIA, HUST Aug 2023 – Jul 2024

Applying Information Bottleneck to Long-Tailed Recognition **Advisor: Professor Shan Tan**

- Developed a novel Balanced Information Bottleneck (BIB) approach for long-tailed visual recognition, achieving substantial improvements in model performance.
- Designed an innovative Mixture of Balanced Information Bottleneck (MBIB) network in an end-to-end fashion, improving recognition accuracy by 1-3% over BIB across widely-used long-tailed datasets.
- Submitted a paper to *Transactions on Machine Learning Research* (TMLR), currently under review.
- Investigated the application of information bottleneck theory to mixture-of-experts (MoE) methods, identifying new opportunities to enhance model efficiency and accuracy.

Research Assistant, HUST AIA Image & Vision Learning Lab, HUST Nov 2022 – Apr 2023

Improving 3D Tooth Segmentation **Advisor: Professor Xiang Xiang**

- Processed, labeled, and analyzed 3D tooth data.
- Engineered and optimized various loss functions to effectively address challenges of imbalanced data.

Competition Experiences

Team Member, Robot Car Competition Feb 2024 – Jun 2024

- Designed and developed a robot car capable of tracking, transporting, and collecting balls by color.
- Placed top-10 in the university competition.

Team Leader, College Robotics Competition Nov 2021 – Jan 2022

- Designed a robot capable of collecting balls using a rotary drum structure inspired by a tennis ball collector.
- Implemented a turntable structure for efficient ball launching.

Selected Course Projects

Developed a Real-Time Bee Detection and Tracking System Apr 2024 – Jun 2024

Course: Pattern Recognition, HUST

- Utilized advanced pattern recognition techniques, including two-stage detection-based tracking, single-stage tracking, and posture-based dense-object tracking, to detect and track bees, which are small, fast-moving, and often arranged in dense swarms.

Built a Public Transportation Itinerary System Jan 2022 – Apr 2022

Course: C Programming, HUST

- Developed a transportation itinerary system in C to track and record the trip records of COVID-19 patients.

Designed a Multifunctional Robot Jul 2023

Course: Deep Learning and Robotics, NUS

- Designed a robot capable of remote-control mobility using Arduino and Raspberry Pi. Developed functionality to recognize different kinds of cats through YOLO.
- Created the advanced model, "Emotion-aware Companion with Helpful Operations (ECHO)," featuring functionality for emotion recognition (CNN), voice chat (ChatGPT API), and object-finding (YOLO and obstacle-avoidance algorithm).

Leadership & Activities

Backbone, Photography Department, School News Center, HUST Sep 2021 – Oct 2023

- Organized and executed events, capturing and producing photos and videos for various occasions, including sporting events, art festivals, and street interviews.
- Organized skills training sessions for group members, including the use of photography equipment and software (e.g., Photoshop, Premiere, After Effects).
- Earned recognition as Outstanding Member.

Team Leader & Small Forward, College Basketball Team of AIA, HUST Apr 2022

- Organized daily training sessions and designed tactics and strategy.

Skills

- Programming Languages: Python, C, Verilog
- Software: 3Done, SolidWorks, Altium Designer, ISE, Multisim, EndNote, LaTeX