Jinwei HU

Tel: +44 07529145856
Email: limboc22@gmail.com
Address: Flat 204, 88 Low Hill, Liverpool, L6 1AT

EDUCATIONAL BACKGROUND

12. 2023- **University of Liverpool**
12. 2027 PhD in Computer Science
10. 2022- **Imperial College London** **Distinction**
10. 2023 MSc in Applied Computational Science and Engineering
09. 2020- **University of Liverpool**
06. 2022 Bachelor of Science with Honours in Computer Science (Artificial Intelligence) **First Class**
09. 2018- **Xi'an Jiaotong-Liverpool University**
06. 2020 Bachelor of Science in Information and Computing Science **First Class**

WORK EXPERIENCE

09. 2024- Present **University of Liverpool** **Research Associate**

- Mainly participate in project "CRoCS: Certified Robust and Scalable Autonomous Operation in Cyber Space" which is funded by Alan Turing Institute (The AI for Cyber Defence (AICD) Research Centre)
- Participate in project "Robustifying Generative AI through Human-Centric Integration of Neural and Symbolic Methods" which is funded by EU Horizon

01. 2024- 01. 2025 **University of Liverpool** **Teaching Assistant**

- COMP338 Computer Vision
- COMP305 Biocomputation
- COMP202 Complexity of Algorithms
- COMP532 Machine Learning and BioInspired Optimisation

03. 2021- 05. 2021 **DXC Technology** **Project Assistant**

- Collected image data for data analysis and automate processes from the back-end.
- Used web scraping libraries to download image data from various online sources.
- Participated in the automation project of online tax paying.
- Positioned the web page element through id attribute, link text including absolute path and relative path, and CSS properties.

07. 2020- 08. 2021 **Alumni MAX** **Project Leader**

- Acted as a project leader in charge of website development and maintenance.
- Beautified and adjusted the front-end webpage and managed the back-end data.
- Continued to improve my proficiency at Java, HTML, and JavaScript.

06. 2020- 09. 2020 **Nanjing Tuoheng UAV System Research Institute** **Project Intern**

- Involved in a project on visual target recognition and tracking method based on deep learning in order to solve the low drone accuracy problem.
- Selected the SSD target detection algorithm based on the CNN network and used the convolution detect and extract different feature maps based on VGG16.
- Installed the development environment required by Tensorflow Object Detection API and tested it on the official Demo, and imported the data for training and testing to build my own model.

PUBLICATION

1st Author **Tapas are free! Training-Free Adaptation of Programmatic Agents via LLM-Guided Program Synthesis in Dynamic Environments**
AAAI Conference on Artificial Intelligence (AAAI 2026 Oral, CCF A · CORE A)*
Falcon: Fine-grained activation manipulation by contrastive orthogonal unalignment for large language model
Annual Conference on Neural Information Processing Systems (Neurips 2025, CCF A · CORE A)*

Explainable AI models for predicting drop coalescence in microfluidics device
Chemical Engineering Journal (JCR Q1 · CAS Q1/Top · IF 13.2)
Enhancing Robustness of LLM-Driven Multi-Agent Systems through Randomized Smoothing
Chinese Journal of Aeronautics (JCR Q1 · CAS Q1/Top · IF 5.7)
Hierarchical testing with rabbit optimization for industrial cyber-physical systems
IEEE Transactions on Industrial Cyber-Physical Systems

Co-Author

Position: Building Guardrails for Large Language Models Requires Systematic Design
International Conference on Machine Learning (ICML 2024, CCF A · CORE A)*
Safeguarding large language models: A survey
Artificial Intelligence Review (JCR Q1 · CAS Q1/Top · IF 13.9)
SIDA: Social Media Image Deepfake Detection, Localization and Explanation with Large Multimodal Model
IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR 2025, CCF A · CORE A)*
Safe Pruning LoRA: Robust Distance-Guided Pruning for Safety Alignment in Adaptation of LLMs
Transactions of the Association for Computational Linguistics (CCF B · JCR Q1 · CAS Q2 · IF 6.9)
Machine learning and physics-driven modelling and simulation of multiphase systems
International Journal of Multiphase Flow (JCR Q1 · CAS Q2 · IF 3.8)
Explainable AI model for predicting equivalent viscous damping in dual frame–wall resilient system
Journal of Building Engineering (JCR Q1 · CAS Q2/Top · IF 7.4)

SCHOLARSHIP

- PhD Full Scholarship funded by the China Scholarship Council (CSC) and University of Liverpool
- ELLIS Manchester Scholarship funded by the University of Manchester

SKILLS AND INTERESTS

- GRE: 155/170 in Verbal Reasoning; 169/170 in Quantitative Reasoning; 3.5/6 in Analytical Writing
- IELTS: 7.0 (7.0 in Listening; 8.5 in Reading; 6.5 in Writing; 6.0 in Speaking)
- Native Chinese speaker and fluent in English
- Personality: vigorous, responsible, inclusive, independent, self-disciplined, optimistic
- Capabilities: strong learning ability, interpersonal communication and teamwork skills, organisational and leadership abilities, well-adaptable, pioneering spirit and an enquiring mind, academic research competence