

# Yuhang YAN (Henry)

 <https://yany-henry.me> |  [yhyan2@cse.cuhk.edu.hk](mailto:yhyan2@cse.cuhk.edu.hk) |  (+86) 18945222225

## EDUCATION

The Chinese University of Hong Kong (CUHK)

Aug. 2021 - Jul. 2025 (expected)

B.Sc. in Computer Science (ELITE Stream)

Hong Kong SAR

- CGPA: **3.739**; Major GPA: **3.803**/4.000 (**Dean's List, Top 10%**)
- **Spring semester exchange** to École Polytechnique Fédérale de Lausanne (**EPFL**), 2023-24
- **Spring semester exchange** to Yuanpei College, Peking University (**PKU**), 2022-23

## RESEARCH EXPERIENCE

**Fact or Fairness? Identifying Over-Balanced Issues** [Slides]

Apr. 2024 - present

Supervised by Prof. Michael R. LYU (CUHK)

Hong Kong SAR

- **Development of the “Fact-or-Fair” Checklist:** Designed the “Fact-or-Fair” checklist using 19 real-world social statistics and three psychological cognitive biases. This tool evaluates large language models (LLMs) and text-to-image (T2I) models with objective and subjective queries to identify biases.
- **Quantitative Metrics for Fact and Fairness Assessment:** Created metrics to quantitatively assess both factual accuracy and fairness in model outputs. Conducted analyses showing the trade-off between content accuracy and equitable representation across demographics.
- **Validation with Different Generative Models:** Conducted experiments with six LLMs and four T2I models to validate the checklist. Results highlighted variations in social bias handling, revealing the different levels of bias mitigation and offering insights for further improving fairness.

**Evaluation on the Vulnerability of Current Generative Models** [Slides]

Feb. 2024 - Jun. 2024

Supervised by Prof. Sabine SÜSSTRUNK (EPFL)

Lausanne, Switzerland

- **“Jailbreak” Analysis & Defense Mechanism Development:** Replicated and tested ten “jailbreak” techniques to evaluate and identify vulnerabilities in generative models. Analyzed weaknesses and adopted effective defenses to enhance models’ security and reliability.
- **Content Moderation and Bypass Detection:** Applied reinforcement learning and greedy search to explore latent space for alternative expressions of sensitive terms. Highlighted limitations of current generative models in preventing malicious exploitation.
- **Bias Detection and Fairness Enhancement in Generative Models:** Identified gender and racial biases in diffusion models and fake-image detectors. Applied fairness enhancement methods with fine-tuning techniques like Fair Diffusion, reducing biases while maintaining model performance.

**Efficient Video Analytics** [Poster]


Jun. 2023 - Sep. 2023

Supervised by Prof. Eric LO (CUHK)

Hong Kong SAR

- Won the **Best Project Award 2023** among 58 undergraduate projects.
- **Multi-modal AI System for Lost-and-Found:** Developed a system using CLIP and OWL-ViT to retrieve data from airport video footage via textual and image inputs. Enhanced AI’s role in lost-and-found services by simplifying searches and reducing manual efforts.
- **Custom Dataset & Model Evaluation:** Built a custom airport video dataset and set evaluation criteria for model performance. Tested and selected top zero-shot object detection and natural language processing (NLP) models, cutting data processing time by 80% and boosting system efficiency.
- **Precise Separation Moment Detection Algorithm:** Designed an algorithm to detect moments when owners and belongings separate. Optimized analysis of video data with high accuracy and recall rate, improving lost-and-found efficiency.

## PUBLICATION

- [1]  Jen-tse Huang, **Yuhang Yan**, Linqi Liu, Yixin Wan, Wenxuan Wang, Kai-Wei Chang, Michael R. Lyu. “*Fact-or-Fair: A Checklist for Behavioral Testing of AI Models on Fairness-Related Queries*.” In submission to the 63rd Annual Meeting of the Association for Computational Linguistics (ACL’25).

## PATENT

- [1] Heng Yang, Yongjie Wu, Tao Long, **Yuhang Yan**. “*A natural language interactive retrieval intelligent security system based on large model*.” *CN118939831A*, 2024, China.







WORK EXPERIENCE

<b>AiMall Technology Co., Ltd.</b> [ <a href="#">Website</a> ] <i>Algorithm Engineer Intern</i>	Jun. 2024 - Aug. 2024 Shenzhen, China
<ul style="list-style-type: none"><li>• <b>Cloud-Based Video and Audio Embedding Storage:</b> Patented a method to embed video and audio frames as vectors in cloud databases, enabling query-based retrieval. This innovation simplifies surveillance, reducing search times from 30+ hours to minutes.</li><li>• <b>Object Detection with Fine-Tuned Vision-Language Models:</b> Built self-trained computer vision algorithms and DeepSORT models for video preprocessing. Used Swift and LoRA to fine-tune VLMs, achieving 95%+ object retrieval success with enhanced matching accuracy.</li><li>• <b>Model Quantization for Efficiency:</b> Optimized LLMs using AWQ and GPTQ techniques, cutting computational needs while preserving performance. Achieved 60% data processing efficiency, enabling faster searches and better system scalability.</li><li>• <b>LLM Self-Recognition Fine-Tuning:</b> Fine-tuned large language models (LLMs) with open datasets like Alpaca and Swift/self-cognition to boost self-recognition. This improved context-aware responses, enhancing system intelligence and interaction quality.</li></ul>	
<b>AIJobTech, Ltd.</b> [ <a href="#">Website</a> ] <i>Co-founder &amp; CTO</i>	Mar. 2024 - present Hong Kong SAR
<ul style="list-style-type: none"><li>• <b>Co-founded AI-Driven Job Matching Startup:</b> Led the creation of a startup automating and personalizing job searches with AI to enhance matching accuracy and efficiency. Secured funding by competing in startup competitions, attracting investors with our innovative approach.</li><li>• <b>Leadership in Development and Partnerships:</b> Directed strategies and coordinated teams to ensure smooth product delivery. Managed developers and designers, fostering collaboration and achieving goals. Acted as technical liaison with partners, enabling seamless B2B integrations.</li><li>• <b>Product Development and AI Innovation:</b> Leveraged LLMs and AI to enhance user experience. Developed features aligning users with suitable roles and enabling quick applications. Implemented personalized resume suggestions based on backgrounds and target positions, boosting success rates.</li></ul>	

ACADEMIC AWARDS AND SCHOLARSHIPS

Dean's List ( <b>Top 10%</b> )	<i>Faculty of Engineering</i>
Outstanding Student Scholarship (15,000HKD), 2022-23 & 2023-24 & 2024-25	<i>S.H. Ho College</i>
Multiple Exchange Scholarship (32,400HKD), 2022-23 & 2023-24	<i>Chinese University of Hong Kong</i>
Best Project Award ( <b>Top 5</b> of 58 UG Summer Research Internships), 2023	<i>Faculty of Engineering</i>
Honorable Mention of Mathematical Contest In Modeling ( <b>MCM</b> ), 2022	<i>COMAP</i>
Matriculation Scholarships for Academic Excellence (5,000HKD), 2021	<i>S.H. Ho College</i>
2 <sup>nd</sup> Prize in National Mathematics Competition for Senior High, 2020	<i>Chinese Mathematics Society</i>
2 <sup>nd</sup> Prize in National Physics Competition for Senior High, 2019	<i>Chinese Physical Society</i>

EXTRA-CURRICULUM ACTIVITIES

 <b>Istria 100 by UTMB</b> <i>21KM trail half-marathon</i>	Apr. 2024 Umag, Croatia
 <b>Hong Kong Mainland Undergraduate Football League</b> <i>Champion; 9 teams from around Hong Kong</i>	Apr. 2024 Hong Kong SAR
 <b>Oxfam Trailwalker 2023</b> <i>Completed the 100KM MacLehose Trail in a continuous 37h25min.</i>	Nov. 2023 Hong Kong SAR
 <b>“INNO Boyue” Cup Mandarin Debate Competition</b> <i>2<sup>nd</sup> runner-up; 24 teams from around the world</i>	Jul. 2022 Virtual
 <b>Fujia “Quanzhou Spring” Volunteer Teaching Project</b> <i>Volunteer teaching for children of incarcerated parents</i>	May 2022 Quanzhou, China
 <b>“Shi'er Jiao” Cup Mandarin Debate Competition</b> <i>Runner-up; 12 teams from around Hong Kong</i>	Apr. 2022 Virtual

SKILLS

<b>Languages</b>	English (Proficient), Mandarin (Native) and Cantonese (Intermediate).
<b>Programming</b>	C/C++, Java, JavaScript, MATLAB, Python{Flask, PyTorch, TensorFlow} and SQL.
<b>Tools</b>	L <sup>A</sup> T <sub>E</sub> X, Conda, Docker, Git, Linux, Spark and Vim.