

Yaowen Ye (Elwin, 叶耀文)

elwin@connect.hku.hk, helloelwin.github.io



RESEARCH INTEREST

I aim to develop responsible and aligned AI that follows human values and interests rather than bringing adverse consequences. Related topics include human-AI collaboration, learning human preferences, red-teaming LLMs, scalable oversight, etc. I am also interested in advancing conceptual research for aligning future super-human intelligence.

EDUCATION

- The Affiliated High School of South China Normal University** 2018.09-2021.6
Major in mathematics, physics, biology, Chinese, English, and politics
- The University of Hong Kong** 2021.09-present
Bachelor of Engineering in Computer Science (GPA: 4.09/4.3)
Machine Learning, Probability and Statistics, Stochastic Processes, Data Structures and Algorithms, etc.

RESEARCH EXPERIENCES

- Research internship at [Yao's Lab](#) at UNC-Chapel Hill** 2023.10-2023.12
Supervised by Dr. Huaxiu Yao. I worked on a project studying the tradeoff of safety alignment in language models and developing alignment techniques that leverage model reasoning capability to reduce over-safety side effects and achieve better balance.
- Research internship at [CoRe Lab](#) at Peking University** 2023.06-2023.09
Supervised by Dr. Yixin Zhu. I worked on a project investigating the heuristic and simulation modes of humans' and deep learning models' physical reasoning. We proposed a new framework that conceptualizes intuitive physics as a dual process. I also studied other cognitive reasoning research topics, including abductive reasoning and concept learning.
- Research internship at [Data Intelligence Lab](#) at HKU** 2021.12-2023.06
Supervisor: Dr. Chao Huang. I worked on projects that leverage generative self-supervised learning to address the data noise and sparsity issues in recommender systems, reducing the performance bias for newly arriving users.

PUBLICATIONS

- "Graph Masked Autoencoder for Sequential Recommendation" SIGIR'23
The 46th International ACM SIGIR Conference on Research and Development in Information Retrieval
Yaowen Ye, Chao Huang, and Lianghao Xia.
- "Masked Graph Transformer for Recommendation" SIGIR'23
The 46th International ACM SIGIR Conference on Research and Development in Information Retrieval
Chaoliu Li, Chao Huang, Lianghao Xia, Xubin Ren, Yaowen Ye, and Yong Xu.

LEADERSHIP

- Co-founder of HKU AI4Good Association** present
A student interest group focused on undergraduate AI research, AI safety, and alignment.
Activities include research seminars, paper reading groups, research mentorship programs,

and invited talks from Ph.D.'s and professors in related fields.

- Co-leader of HKU Astar (former HerKules Robomaster-AI) Team** 2022-2023
A team focused on RoboMaster competitions. I led the vision team and did technical work for
(1) Auto-aiming: Designing algorithms to detect enemy robots and control the gun to attack.
(2) Localization: Locating robots in the point cloud space with RGBD cameras.
- Captain of Lee Shau Kee Hall Choir** 2022-2023
A cultural team at HKU Lee Shau Kee Hall. I led the LSK Choir throughout the year, organized weekly practice, and successfully participated in the 2023 Interhall Choir Competition.

AWARDS

- Silver Award** 2022.10
The 1st EEG & AI Competition organized by the Faculty of Education of HKU (HK\$ 2,000)
- Young Tsun Dart Scholarship** 2022-23
Merit-based Scholarship for students in HKU Engineering faculty (HK\$ 15,000)
- Noel Chau Scholarship** 2021-22
Merit-based Scholarship for HKU students (HK\$ 20,000)
- Dean's Honours List** 2021-22
Merit-based Award for students in HKU Engineering faculty
- Ho Fook and Chan Kai Ming Prize** 2021-22
Scholarship awarded by HKU (HK\$ 5,000)
- Outstanding Student of the Year** 2019.12
Science Talent Program jointly organized by China Association for Science & Technology and the Ministry of Education of the PRC
- Outstanding Student** 2019.08
Tsinghua University Global Innovation Exchange Summer Camp
- First Prize** 2018.01
National Olympiad in Informatics in Guangdong province organized by CCF

ACTIVITIES

- Talk on *Machine Learning Introduction* at Dong Yuan High School, Guangdong 2022.06
- HKU BrainStem invited sharing on research experience 2023.02
- Serve as a reviewer for *BrainStem Scientific Journal* 2023.08
- CSSAUD invited sharing on pursuing undergraduate research 2023.09

SKILLS

Python, C++, Haskell, R; Data structures and algorithms;
Machine Learning and Deep Learning; PyTorch and TensorFlow.
Adobe Photoshop, Final Cut Pro, Microsoft Office.