

# Wenbin Hu

Incoming MPhil Student  
Department of Computer Science and Engineering  
Hong Kong University of Science and Technology, HK

+852-63405208  
✉ whuak@connect.ust.hk  
🌐 GitHub Profile  
🌐 LinkedIn Profile

## EDUCATION

---

### Hong Kong University of Science and Technology, HK

Sept 2020- Jul 2024

*B.Eng in Computer Science*

- **Awards:** Dean's List (Term GPA > 3.7, Top 10 Percentile).
- **AI Related Courses:** Machine Learning, Deep Learning, Computer Vision, Natural Language Processing, Graph Machine Learning, Reinforcement Learning.
- **CS Related Courses:** C++, OOP, Computer Architecture, Operation System, Algorithm, Software Engineering, Computer Graphics, Web Search Engine.
- **MATH Related Courses:** Multivariable Calculus, Linear Algebra, Abstract Algebra, Probability Theory, Convex Optimization, Mathematic Analysis, Ordinary Differential Equation.

### Technical University of Munich, Munich

Apr 2023 - Aug 2023

*Exchange Programme in TUM Informatics*

- **Awards:** Graded the highest level: 'Very Good'
- **Seminar Talk:** I have made a 40-minute presentation about generative NeRF in a graduate level seminar.

### Fudan University, Shanghai

Jun 2022 - Aug 2022

*Summer Exchange Programme*

- **Research Topic:** Probabilistic Graph Model: Bayesian Inference and Learning.

## PUBLICATIONS

---

- **Node Level Graph Autoencoder: Unified Pretraining for Textual Graph Learning**, Wenbin Hu\*, Huihao Jing\*, Qi Hu\*, Haoran Li, Yangqiu Song. Arxiv preprint. [\[paper\]](#)
- **Mitigating the Alignment Tax of RLHF**, Yong Lin, Hangyu Lin, Wei Xiong, Shizhe Diao, Jianmeng Liu, Jipeng Zhang, Rui Pan, Haoxiang Wang, Wenbin Hu, Hanning Zhang, Hanze Dong, Renjie Pi, Han Zhao, Nan Jiang, Heng Ji, Yuan Yao, Tong Zhang. To appear in EMNLP 2024 [\[paper\]](#).
- **Attacking by Aligning: Clean-Label Backdoor Attacks on Object Detection**, Yize Cheng\*, Wenbin Hu\*, Minhao Cheng. Arxiv preprint. [\[paper\]](#)

(\* represents equal contribution.)

## RESEARCH INTERESTS

---

My research interest is centered around **LLM Reasoning and Safety**.

## SKILLS

---

**Programming Language:** C/C++, Python

**Library/Framework:** Pytorch, NumPy, OpenGL.

**Tools:** Latex, Markdown, Git, Microsoft Office.

**Language:** Chinese (Native), English (Fluent), Cantonese (Good).

**Soft Skill:** Academic Writing, Communication, and Presentation; Strong Self-motivation and Self-learning Ability.

## PERSONAL INTERESTS

---

Piano, Basketball, Running, Exploring New Things.