HAO LIU

Room 128B, Building Zijing 1#, Tsinghua University, Beijing, 100084, China
liu-hao20@mails.tsinghua.edu.cn | (+86) 15510851552

EDUCATION

Tsinghua University (THU) Sep 2020– June 2024
Bachelor of Engineering (Computer Science and Technology)
- Computer Science Major GPA: 3.3 /4.0
- Core Courses: Fundamentals of Programming: 4.0
  Discrete Mathematics:3.6
  Linear Algebra: 3.6
  Foundation of Object-Oriented Programming: 4.0
  Artificial Neural Networks: 4.0
  Advances in Autonomous Driving and Intelligent Vehicles: 4.0 (A+)
  English Reading and Writing for academic purpose: 4.0 (A)
  English listening and speaking for academic purpose: 4.0 (A+)

New York University (NYU) Sep 2024– Pesent
Master of Science (Cognitive Psychology)

Beijing Institute for General Artificial Intelligence (BIGAI) Sep 2022– June 2024
Intern (Multiagent Group)

RESEARCH EXPERIENCE

ICLR 2025 in submission: BlockFound: CUSTOMIZED BLOCKCHAIN FOUNDATION MODEL FOR ANOMALY DETECTION (https://arxiv.org/pdf/2410.04039)
Advisor: Xinyu Xing / Northwestern (Remote) March 2024– June 2024
- Help build and test LLM based anomaly detection algorithm on blockchain transaction.
- Carefully clean and analyze the transaction data, including developing multi-thread algorithms to efficiently preprocess the data.
- Train and test different LLM models on fitting the dataset, LoRA for example.

NeurIPS 2024 Benchmark Track, Poster: AdaSociety: An Adaptive Environment with Social Structures for Multi-Agent Decision-Making (https://neurips.cc/virtual/2024/poster/97511)
Advisor: Siyuan Qi, Yaodong Yang, Xue Feng, Songchun Zhu / Peking University June 2023– June 2024
- Built a Grid world Based environment for studying generalized social behavior. The main dynamic of the environment consists of 3 parts: resource, event and agent. Agent could gather resources to make higher products though engaging events.
- Using RLLib to deploy classic algorithms like MAPPO, MADDPG to see the result.

Watch-Talk-Help: Multiagent communication based on Large Language Model and Theory of Mind
Advisor: Tianmin Shu, Joshua Tenenbaum / MIT June 2023– September 2023
- Made further improvements on the basis of previous two articles of Prof. Tenenbaum, watch-and-help (https://arxiv.org/pdf/2010.09890.pdf)and online-watch-and-help (https://arxiv.org/pdf/2301.05223.pdf). Added natural language communication generated based on the Large Language Model between the human agent and the robot agent.
- Based on this natural language interaction, I defined rules for updating the agent's belief and goal (Theory of Mind). We borrowed ideas from Prof. Tenenbaum's previous Too many Cooks（https://arxiv.org/abs/2003.11778） And Bayesian Theory of Mind（ http://web.mit.edu/9.s915/www/classes/theoryOfMind.pdf）.
- My work tried to determine when to engage in natural language communication by calculating the possibility differences between the actions induced by the combined beliefs of human mind and robot mind, and the actions induced by beliefs only in human mind. The beliefs of two agents and their plans will gradually converge as they communicate, ultimately forming a joint planner to efficiently complete the task together.

Pairwise learning / Functional embedding
Advisor: Zhiting Hu / UCSD(Remote) June 2022– September 2022
- Come up with a general framework to unify data & parameter space by learning pairs of data.
- Experiment on different architectures like CNNs & Transformers and different areas like CV & NLP

Stock prediction based on large scale pretraining model February 2022–September 2022
Advisor: Maosong Sun / Tsinghua University
- Designed a complete finetuning pipeline based on Macro, Meso and Micro views of the stock market.
- Pretrained a transformer model on the large Chinese financial corpus using Pytorch.
- Wrote web crawlers to get more than 15G of Chinese financial text, built these data into a large pretrain corpus.

A Self-supervised pretraining model based on Encoder-Decoder-Unmasker architecture December 2021–February 2022
Advisor: None
- Developed a complete self-supervised model based on an Encoder-Decoder-Unmasker model training on ImageNet1K.
- Used the Encoder as the backbone net for image classification task and get competitive result. (Need further experiments)
- Explained the model intuitively using causal representation learning.
- Hand reproduced ViT and use it as the base architecture for my model.

A Q-A model based on query mechanism September 2021–November 2022

Advisor: Minlie Huang / Tsinghua University

- Built the Query mechanism in two different ways. One use traditional BM25 similarity algorithm, one pretrained on BERT to predict the correlation degree between a pair of Q-A.
- Query the corpus while pretraining on the LongLM (A pretrained transformer-based model on Chinese corpus).
- Inferenced the model using test set and got better result than baseline. (Both BLEU and ROUGE score).
- Cleaned 1000K+ raw QA pairs from Zhihu and build the corpus.

Pollution Image classification system December 2017–October 2019(High school)

Advisor: Mingming Cheng / Nankai University

- Built a simple supervised learning classification model using SVM algorithm.
- Extracted features by hand using Dark channel algorithm, information entropy and Color Histogram in HSV color space.
- Wrote web crawlers to get more than 10K of different kinds of pollution image.

PAPERS

- **ICLR 2025 in submission: BlockFound: CUSTOMIZED BLOCKCHAIN FOUNDATION MODEL FOR ANOMALY DETECTION (https://arxiv.org/pdf/2410.04039)**

- **NeurIPS 2024 Benchmark Track, Poster: AdaSociety: An Adaptive Environment with Social Structures for Multi-Agent Decision-Making (https://neurips.cc/virtual/2024/poster/97511)**

COURSE EXPERIENCE

- Open course: Stanford CS231n, Computer Vision (Feifei Li) September 2020–December 2020
- Open course: Stanford CS224n/Ling284, Natural language processing with deep learning (Chris Manning) January 2022–February 2022
- Open course: Introduction to reinforcement learning (David Silver) May 2021– August 2021
- Open course: Stanford CS224w, Graph neural networks (Jure Leskovec) May 2021– August 2021
- Open course: UCB CS285, Deep reinforcement learning (Sergey Levine) May 2021– August 2021
- Open course: Next Step of Machine Learning (Hongyi Lee) January 2021– February 2021
- Open course: Geometric Deep Learning (Michael M. Bronstein) June 2022 – September2022
- Open course: Stanford CS330 Deep Multi-Task and Meta Learning (Chelsea Finn) September2022 – December 2022
- Open course: Categories for AI (DeepMind) March 2023 – December 2023

ACADAMIC CONFERENCE/LECRTURE ATTENDENCE EXPERIENCE

China Theory Week 2018 | Tsinghua University | Guest high school student September 2018

- Topic: Theoretical computer science
- Hosted by Andrew Yao, lecture by Seth Pettie, Kasper Green Larsen etc.

World Intelligence Congress |Tianjin University | Guest high school student May 2019

- Topic: Conscious Turing Machine
- Lecture by Manuel Blum

2021 BAAI Conference|Beijing | Guest June 2021

- Topic: Artificial Intelligence
- Lecture by Yoshua Bengio

AWARDS

YINGCAIJIHUA (Research training for science and technology innovation for high school students) October 2019

- Outstanding paper award
- Outstanding Student award

"21st Century Cup" National English Speaking Competition October 2018

- Second award (Tianjin)

SKILLS

Language: Mandarin Chinese (First language), Fluent in English
Programming: C, C++, Python (Pytorch, Gym, Ray(RLlib)), HTML, FPGA
Applications: Matlab, WordPress