

Mingjia Huo

📞 217-953-1827 | ✉️ mhuo@ucsd.edu | 🌐 mignonjia.github.io/

Education

University of California, San Diego (UCSD)

PhD in Electrical and Computer Engineering (GPA: 4.0)

- Advisor: Hao Zhang

California, US

Sep 2023 - now

University of Illinois, Urbana-Champaign (UIUC)

MS in Electrical and Computer Engineering (GPA: 3.95)

- Advisor: Kirill Levchenko

Illinois, US

Aug 2020 - Dec 2022

Peking University

BS in Computer Science (Turing Class, GPA: 3.76, rank top 10%)

- Advisor: Qun Huang

Beijing, China

Sep 2016 - Jun 2020

Research Projects

Evaluating LLMs on Games

Feb 2025 - now

- Evaluated the performance of large language models on diverse games including Sokoban, Super Mario, Tetris, 2048, and Candy Crush.
- Developed LLM agents with visual and memory components to interact with game environments.
- Trained LLMs with reinforcement learning to enhance performance on gameplay tasks.
- Open-sourced the benchmark and agent framework: <https://github.com/lmgame-org/GamingAgent>.

LLM Watermarking

Sep 2023 - Feb 2024

- Applied LLM watermarks by adjusting LLM logits during inference time, and trained networks to adjust watermark strength based on previous tokens.
- Designed a multi-objective optimization framework to balance detectability and semantic coherence.
- Applied Gumbel-Softmax and a straight-through estimator to preserve gradients.
- Evaluated on C4 realnewslike dataset and showed our method significantly improved the Pareto frontier of detection-semantics trade-off curves.
- Explained the rationality of learned parameters with respect to part-of-speech (POS) tags.

Multi-Modal Large Language Model for Protein Function Prediction

Jul 2023 - Oct 2024

- Applied multimodal learning (LLAVA) to perform instruction tuning based on Llama2-13B on one million QA data points.
- Utilized Pytorch Distributed Data Parallel (DDP) for multi-GPU training.
- Evaluated the performance on open-text generations and classification tasks using F1-score, perplexity, BLEU, and SimCSE.
- Visualized the learned embeddings using t-SNE.

Publication

[1] Mingjia Huo, Sai Ashish Somayajula, Youwei Liang, et al. Token-Specific Watermarking with Enhanced Detectability and Semantic Coherence for Large Language Models. International Conference on Machine Learning (ICML), 2024.

[2] Mingjia Huo, Han Guo, Xingyi Cheng, et al. Multi-Modal Large Language Model Enables Protein Function Prediction. (In preparation)

[4] Mingjia Huo, Maxwell Bland, Kirill Levchenko. All Eyes On Me: Inside Trackers' Exfiltration of PHI from Healthcare Providers' Online Systems. Proceedings of the 21th ACM Workshop on Privacy in the Electronic Society (WPES), 2022.

Work Experience

Trova AI, Inc.

Illinois, US

AI Software Development Intern

May 2022 - Aug 2022

- Conducted customer segmentation using clustering methods for Snap-on, a US manufacturing company.
- Performed feature engineering on the purchase history of 6 million customers spanning from 2010 to 2022.
- Trained XGBoost model to predict individual purchase intention, and improved F1-score by 14%.
- Presented findings to the company's leadership and the franchisee training sessions for deployment.

Biomap

Beijing, China

Machine Learning Engineer Intern

Jul 2023 - Sep 2023

- Leveraged xTrimoPGLM-1B as the encoder to extract protein embeddings from amino-acid sequences.
- Trained a lightweight adapter to map protein embedding to the embedding space of Llama2.

Skills

Tool PyTorch, Matlab, Kubeflow, Snowflake, Adobe Illustrator

Programming Python(Fluent), SQL(Fluent), C, C++

University Work Experience

2025 Teaching Assistant, ECE 268: Security of Hardware Embedded Systems

UCSD

2022 Teaching Assistant, CS 461: Computer Security I

UIUC

2022 Teaching Assistant, ECE 445: Senior Design Laboratory

UIUC

Selected Awards

2019 **Fellowship**, Hui-Chun Chin and Tsung-Dao Lee Chinese Undergrad Research Endowment

2015 **Silver Medal**, Chinese Mathematical Olympiad

2015 **Gold Medal**, Chinese Girls' Mathematical Olympiad