Jingyuan He

jingyuan_he@outlook.com • (858) 699-1238 • https://j7he03.wixsite.com/jingyuan-he

EDUCATION

Carnegie Mellon University Pittsburgh

Master of Science in Intelligent Information Systems | GPA: 4.00/4.33 August 2023 - May 2025

Selected Coursework: Search Engine, Large Language Models, Machine Learning, Natural Language Processing, Multimodal Machine Learning

University of California, San Diego San Diego

Bachelor of Science in Mathematics-Computer Science | GPA: 3.94/4.0 August 2020 - June 2023

Minor in Computer Engineering | HKN-IEEE Member | Grader of Math184: Enumerative Combinatorics

Selected Coursework: Operating Systems, Deep Learning, Recommender Systems, Probability and Stochastic Process, Graph Theory, Statistics

SKILLS

Programming Languages: Python, Java, SQL, R, HTML, CSS; Intermediate: C++, C, MATLAB, JavaScript, Assembly, SystemVerilog

Tools: PyTorch, Huggingface, Scikit-learn, OpenCV, Linux, Android Studio, Unity, TensorFlow, AWS, Slurm

EXPERIENCE

Meta Menlo Park, California, United States

Software Engineer 06/23/2025 – present

- Work on content understanding for advertisement recommendation with Large Language Models.

Meta Menlo Park, California, United States

Software Engineer Intern 05/13/2024 – 08/02/2024

- Reproduced and compared ID-based, content-based recommenders, ***HSTU*** and ***TASTE***, on AmazonReviews datasets and industrial data.
- Scaled up TASTE training on ***T5***, ***Pythia*** and ***Llama3-8B*** with distributed training technique ***(DDP and FSDP)***.

The Yiddish Arts and Academics Association of North America San Diego, United States

Web Developer Intern 03/28/2022 – 06/27/2022

- Maintain Wordpress-based websites in an efficient and easy-to-navigate manner by front-end plugins, HTML, CSS, and JavaScript.

Guangxi Communication Planning and Design Consulting Corporation Nanning, China

Technical Intern 08/30/2021 – 09/03/2021

- Construct program to remove redundant tasks in management system based on semantic similarity with public ***word-and-phrases-divisor***.
- Identified boundaries of target location entities by public ***BaiduMap API*** system and store into ***mySQL*** database.

PROJECTS

Open Recommendation Benchmark for Reproducible Research with Hidden Tests CMU, Pittsburgh, United States

Research Assistant February 2025 – June 2025

- Designed and developed ***ORBIT***, a unified sequential recommendation benchmark that enables fair and reproducible model comparison**.**
- Built ***ClueWeb-Reco***, a high quality and privacy-preserving web recommendation dataset derived from mapping private collected U.S. browsing sequences to public webpages through a semantic matching process on large-scale dense retrieval.

Efficient Dense Retrieval with Boundary-Aware Cluster Routing CMU, Pittsburgh, United States

Research Assistant September 2023 – May 2024

- Developed ***BACR***, a model to predict the cluster to perform fine-grained search in pre-built cluster-based index Approximate Nearest Neighbor Search (ANNS) that improve ANNS recall by 8.7% compared to ***FAISS*** IVFFlatIP on MSMARCO WEB SEARCH.
- Worked with ***DiskANN***, a memory-SSD hybrid ANN index and implement a multi-node index construction function.

Mixture of Experts Neuro-Computational Network on Fusiform Face Area The Cottrell Lab, UCSD, San Diego, United States

Undergraduate Research Assistant January 2023 - June 2023

- Trained a network of word, object, and face experts to confirm subordinate-level image classification capacity of Fusiform Face Area.
- Configured different network architecture on ***ResNet-18***, ***custom CNN***, ***gating layers*** and achieved 99.4% expert gating accuracy.
- Generated, processed large datasets with ***OpenCV*** and speeded up data-loading from 1 hour to 13 minute per epoch.

Question Answering System CMU, October 2023 – December 2023

- Developed a rule-based polar question detector to prompt ***Mistral-7B*** to answer both factual and binary questions in a 4-people team.
- Fine-tuned ***t5-base*** on Wikipedia datasets for question generation and incorporated ***sentence-t5*** for retrieval augmented answer generation.

Transformer-Based User Intent Classification on Amazon Massive Intent Dataset UCSD, December 2022

- Fine-tuned ***bert-base-uncase*** model on 60 classes of user intent to achieve an 87.32% sentence-level classification in a 5-people team.
- Utilized ***UMAP*** to compare clustering on supervised contrastive learning loss, cross-entropy loss, and simple contrastive learning loss.

Image Captioning with Convolutional Neural Networks and LSTM UCSD, November 2022

- Implemented an encoder-decoder network consists of a ***custom CNN*** or modified ***ResNet-50*** followed by a ***LSTM***, to generate captions for *MSCOCO* dataset in a 3-people team. Achieve an average 58.04% BLEU1 score and 86.53%-100% BLEU4 scores for the best captions.

ZooSeeker Android Software Development UCSD, March 2022 - June 2022

- Examined ***Agile Management*** and utilized ***Android Studio*** to co-develop application ZooSeeker to offer navigation of San Diego Zoo.

PUBLICATIONS

He, J., Liu, J., Oberoi, V. V., Wu, B., Patel, M. J., Mao, K., ... & Xiong, C. ORBIT-Open Recommendation Benchmark for Reproducible Research with Hidden Tests. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems Datasets and Benchmarks Track*.

Coelho, J., Ning, J., He, J., Mao, K., Paladugu, A., Setlur, P., ... & Xiong, C. (2025). Deepresearchgym: A free, transparent, and reproducible evaluation sandbox for deep research. arXiv preprint arXiv:2505.19253.

Wang, T., He, J., & Xiong, C. (2024, November). RAGViz: Diagnose and Visualize Retrieval-Augmented Generation. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: System Demonstrations* (pp. 320-327).