Alex Fang

linkedin.com/in/spaghectic/ | afangster.github.io | (469) 592-5610 | ahtfang@gmail.com

EDUCATION

Carnegie Mellon University | *Masters of Science (ML/NLP @ SCS LTI). GPA: 4.0/4.0*   Pittsburgh, PA | December 2025
- **Coursework (In Progress):** Generative AI, Advanced Natural Language Processing, Introduction to Machine Learning
- **Activities:** NeuLab @ CMU

University of California, Los Angeles | *Bachelor of Science in Computer Science. GPA: 3.8/4.0*   Los Angeles, CA | June 2024
- **Coursework**: Neural Networks and Deep Learning, NLP, Machine Learning, Systems and Signals, Linear Algebra, Discrete Math, Probability and Statistics.
- **Activities:** Tau Beta Pi, Wong Research Lab, Association of Computing Machinery.

WORK EXPERIENCE

Murata Electronics | *Software Engineering Intern*   Carrollton, TX | June 2023 - September 2023
- Designed and implemented an Expo React Native service application prototype which enables a carrier agnostic route to connect cellular IoT devices to cloud services via Azure IoT Hub.
- Aimed to facilitate cellular IoT end device application development and reduce time-to-market for clients.
- Visualized device JSON data routed through Azure IoT Hub event groups and controlled/configured a device from app using direct methods/device twin's desired properties using a Node.js and Express.js backend.

PROJECTS

End-to-End NLP System | *CMU*   Pittsburgh, PA | October 2024
- Developed an end-to-end NLP system using a Retrieval-Augmented Generation (RAG) system with FAISS and BM25 vector retrieval methods, implemented using LangChain, and roberta, Mistral, Llama, and a LoRA-finetuned Llama LLMs.
- Generated a fine-tuning dataset and testing dataset by generating question-answer pairs with context from a document dataset, compiled through web-scraping.
- Achieved 68% F1-score, 49% exact match, 67% precision, and 70% recall with the LoRA-finetuned Llama model using FAISS retrieval on the generated test dataset, showing significant improvement over Llama without RAG.

Factuality and Fairness in LLMs | *UCLA*   Los Angeles, CA | January 2024 - March 2024
- Investigated three open-source LLMs (Phi-2, Mistral, Gemma) leveraging zero-shot prompting and in-context learning to evaluate fairness/factuality accuracy and f1-score from claims in the UniLC benchmark dataset.
- Developed and fine-tuned custom prompts through prompt engineering to optimize model performance across different prompting techniques and models, leading to a 11% and 14% improvement in accuracy and F1-score respectively.
- Achieved 78% accuracy and 0.75 F1-score on testing dataset using the Mistral-Instruct model to generate evidence for zero-evidence-evaluation prompting on the Phi-2 model to generate predictions.

Deep Learning Models for EEG Classification | *UCLA*   Los Angeles, CA | January 2024 - March 2024
- Investigated and designed CNN, LSTM, CNN+RNN, and CNN+LSTM architectures using PyTorch and data preprocessing techniques (data cropping) on classification of 4-class motor imagery EEG signals (BCI Competition IV, Dataset 2a).
- Achieved ~73% classification accuracy on all-subject testing with the CNN model and ~68% classification accuracy on subject-wise testing with the CNN+LSTM model.

RESEARCH

Behavior Box | *Research Assistant under Dr. Graham Neubig*   Pittsburgh, PA | September 2024 – Present
- Automatically identify types of text that language models underperform/over-perform (relative to other models) on, and describe these low-level "slices" in a human-interpretable way.
- Implemented a data visualization pipeline to investigate performance differences between Llama models on the Dolma dataset.

Laboratory Research at UCLA | *Research Assistant*   Los Angeles, CA | April 2022 – January 2024
- Evaluated single-cell image segmentation performance of a deep-learning model against human-generated (ground truth) and MATLAB model generated segmentations with DICE dissimilarity scores and IoU metrics.
- Achieved 0.001 average DICE score and 0.998 average IoU score between deep-learning model and ground truth.
- Facilitated further research by improving image segmentation accuracy on clustered images with overlapping cells.

SKILLS

Languages and Frameworks: Python, C++, C, Bash, PyTorch, NumPy, LangChain.

Tools: HuggingFace, Microsoft Azure, Google Cloud Platform, Git, Pandas, Matplotlib, Seaborn.