Zile Huang

zh3083@nyu.edu | linkedin.com/in/llleeeoooh | https://github.com/powerleo101 | leoh.us

An electronic version of this CV with hyperlinks: www.leoh.us/CV.pdf

ABOUT

I am a first-year student passionate about Computer Science and Philosophy. I am eager to learn new stuff, and, believing in Feynman's philosophy of "what I cannot create, I do not understand", I spend a lot of my free time doing fun projects.

EDUCATION

College of Art and Science, New York University ***Expected Graduation: May 2028***

Bachelors of Arts in Computer Science and Philosophy | GPA 4.0/4.0

Coursework: Computer System Organization, Operating System, Applied Internet Technology

Kent School ***Graduated: May 2024***

High School Diploma, Cum Laude | GPA 3.94/4.0

SKILLS

Languages JavaScript(1yrs) · Python(3yrs) · C/C++ (5yrs) · Lua (2yrs)

Software Docker · PyTorch · TensorFlow · NumPy · SciPy · HuggingFace Family · VLLM · Pandas · Jupyter · Git · Bash

EXPERIENCE

Computer Science Data Annotation ***May 2024 - Present***

DataAnnotation.Tech *New York, NY*

- Annotated hundreds of data points for Large Language Model (LLM) training for highly technical CS questions.
- Prompt engineered difficult questions for LLMs for potential unfactual or faulty response, and provided corrections.
- Contributed to creating a LLM training dataset for improving models' intelligence in CS.

PROJECTS

Mini Deep Learning Framework (In-progress, ~40 hours) - Github ***Dec 2024 - Present***

- Learning and building a deep learning framework capable of training and inference of deep neural networks with a style similar to George Hotz's TinyGrad.
- Framework is capable of define-by-run model declaration, automatic differentiation, lazy computation, intermediate representation of ML logic, etc., and is extensible to have hardware acceleration, runtime optimization, etc.

Classification of Lumbar Spine Degeneration from MRI Scans (~100 hours) - Github Competition ***Jun 2024 - Aug 2024***

- On a dataset of ~2000 patients' spinal MRI scans, trained an ensemble of two stage classifiers that performs segmentation on regions of interest (5 disc levels) and then classification, and another ensemble of 3d classifiers that utilizes attentions for regions of interest, whose weights are modified from 2d pretrained models.
- Received silver medal (ranked 78 out of 1800+ teams) in the competition held by RSNA on Kaggle.

Classification of Harmful Brain Activity Using EEG Signals *(~100 hours)* - Competition ***Feb 2024 - Apr 2024***

- Converted time-series EEG signals from ~2000 patients over 17k+ sessions to spectrogram with data augmentations.
- Trained an ensemble of 10+ vision/waveform hybrid classifiers for 6 abnormal brain activities on EEG signals and their spectrograms. Provided insights for research in multi modal approaches in classifying EEG signals.
- Received silver medal (ranked 64 out of 2700+ teams) in the competition held by Harvard Medical School on Kaggle.

Weakly-Supervised Detection of Dementia from MRI scans *(~50 hours)* - Writeup ***Jul 2023 - Aug 2023***

- Researched on detection of dementia from MRI scans with Dr. Cross-Zamirski.
- Experimented with weakly supervised training of vision transformers with patient's metadata integration. Achieved state-of-the-art accuracy on the OASIS dataset containing 400+ patients .
- Writeup published in Horizon Academic Journal

Activities

Competitive Programming - Self-taught since high school. USACO Platinum. Codeforces 1600+. ***Sep 2021 - Present***

Google Cloud Next '24 Kaggle Hackathon - Second place at the hackathon held at Google Cloud 24 Conference, trained top performing LLM to play trivia questions, earning 7000$ in total prize money ***Apr 2024***

Kaggle Competition Expert - Participated in machine learning competitions. Top 2000 out of all (20k) Kaggle competitors. ***Nov 2024***