YUNWEI ZHAO

New York, NY | 206-351-1702 | yz2959@cornell.edu | LinkedIn | GitHub

EDUCATION

Cornell Tech (Cornell University), New York, NY — Expected Graduation: May 2024
Master of Engineering - Computer Science
Ongoing course: HCI and Design, Machine Learning Engineering, Business Fundamentals

University of Washington, Seattle, WA — Sep 2019 - June 2023
Bachelor of Science - Computer Science | GPA: 3.84/4.0 (Top 7% of the Class)

- **Relevant Coursework**: Artificial Intelligence, Natural Language Processing, Deep Learning, Computer Vision, Autonomous Robotics, Computational Biology, Database Systems, Distributed Systems

RESEARCH EXPERIENCE

H2Lab @ UW - Research Assistant - Mentor: Professor Hannaneh Hajishirzi — Sep 2022 - June 2023
Research Engineer on the topic of fine-tuning and evaluating large-language models with various prompts.

- Explored the correlations between different prompts and pre-trained language models (PLMs) on natural language understanding and generation tasks using **Python**, **Hugging Face**, and **PyTorch**.
- Fine-tuned 15 large language models across several Nvidia A100s and RTX 6Ks via CUDA, utilizing 27 handcrafted prompts across tasks: entailment (MNLI, RTE, CB), multiple choice QA (BoolQ, MultiRC), and commonsense reasoning (WSC, COPA, WiC).
- Evaluated the performance of different prompt-PLM combinations using four metrics: Pearson, Spearman's rank, Kendall rank, and Hellinger distance.
- Collaborated with a team of researchers, communicated research findings weekly, and documented results in a comprehensive report.

ICTD Lab @ UW - Research Assistant – Mentor: Professor Richard Anderson - [Paper] [App on Google Play] — Sep 2022 - Dec 2022
eKichabi, a Tanzania farming directory app to enable users to search for local businesses.

- Paper under review at ACM CHI Conference 2024.
- Leveraged survey data and usage log analysis using **Python**, **Pandas**, and **NumPy** to dissect user patterns and behaviors. Uncovered pivotal insights, such as USSD vs. Android app usage trends, age-based user preferences, and the seasonal impact on app usage.
- Mastered the Android app's codebase and, over 32 iterative versions, introduced user authentication and established efficient logging systems using **Java** in Android Studio in an agile environment.
- Optimized the app's backend by merging the Android server with the USSD server hosted on PythonAnywhere using **Python**, working closely with Django and Database Engineers.
- Achieved a 25% reduction in the app's storage footprint, from 29.52MB to 21.94MB, by re-organizing codebase, improving code logic, and optimizing the sizing of images. Addition results included resolving critical memory and null pointer exception issues.
- Conducted hands-on UX testing with various groups, including farmers and business owners across 300 villages in Tanzania.

Wang Lab @ UW - Research Intern - Mentor: Professor Sheng Wang - [Paper] [GitHub] [Doc] [News] — June 2022 - Sep 2022
biotranslator, an open-source software that translates textual description to non-textual biological data (e.g., gene expression).

- Paper "Multilingual translation for zero-shot biomedical classification using BioTranslator" published on Nature Communication.
- Designed and built open-source cross-modal translation software and API using **Python**, **PyTorch**, and **PyPI/TestPyPI**.
- Refined API code by 65% (~1300 lines), streamlining the user experience and data input methods.
- Introduced a feature to annotate user protein data with textual embeddings, bolstering the software's utility.

Crit2SQL, a novel, large-scale Criteria2SQL dataset that allows users to translate clinical trial criteria into SQL queries.

- Paper under ACL Rolling Review.
- Analyzed and corrected errors in SQL queries by OpenAI Codex
- Annotated over 400 natural language biomedical queries with MySQL in one week.

Wang Lab @ UW - Research Assistant - Mentor: Professor Sheng Wang - [Paper] — Mar 2022 - June 2022
Gemini, a novel network integration method that uses memory-efficient high-order pooling to represent and weight each network according to its uniqueness.

- Paper "Gemini: memory-efficient integration of hundreds of gene networks with high-order pooling" published on Bioinformatics.
- Migrated the Diffusion Component Analysis (DCA) framework from **MATLAB** to **Python.**
- Conducted baseline model testing using various datasets and developed and implemented evaluation metrics.

INDUSTRY EXPERIENCE

Ketogenic.com - Software Engineer Intern - Flutter, Dart, WebSocket, Firebase - [Web] — Dec 2023 - Present
Multi-platform App Developer for Pro2Col, a healthcare mobile app designed to offer professional health advice and planning tools, enabling users to track daily progress, compete and socialize with peers, and share their achievements.

- Implemented key UI screens including workout main/start screens (both portrait and landscape modes), leaderboard, sign-up flow, health and home screens, along with multiple health sub-screens, nutrition pages, and tracker features for steps, fasting, and water using **Flutter** & **Dart**.
- Engineered a robust chat feature using **WebSocket** and **Firebase**, supporting cross-device personal and group communications. Integrated multimedia sharing capabilities (text, image, video), in-app media capture, and voice recording functionalities.
- Pioneered API design for new features, particularly within the leaderboard module. Expertly defined API output formats tailored to design requirements, collaborated with backend teams for API implementation, ensuring seamless integration and functionality.
- Executed the addition of multiple dynamic features across the app, including a nutrition search and results display, food item details, enhancing the overall user experience and app functionality.

AeroSpec - Software Developer - Flutter, Dart - [News] [Web] Mar 2021 - Dec 2021

Multi-platform App Developer for AeroSpec, a real-time sensor visualization tool that enables users to monitor their exposure to air pollutants and track Air Quality Index (AQI) levels.

- Implemented responsive and adaptive designs ensuring compatibility across diverse devices. Ensured the system's robustness in handling real-time data streams from Aero, our all-in-one sensor platform.
- Collaborated closely with 2 designers, 1 product manager, and 3 engineers to ensure accurate AQI representation and playback, enhancing user comprehension and experience.
- Improved the user interface and experience through iterative research, Alpha Testing, and user interviews, ensuring the platform met user needs and expectations.
- Constructed 31 reusable UI components using **Flutter** & **Dart**, streamlining the development process and ensuring consistency across the platform.

LEADERSHIP AND SERVICES

Teaching Assistant - CSE473: Introduction to Artificial Intelligence - Professor Luke Zettlemoyer - University of Washington Fall'22

- Actively engaged in course development, answered questions on Ed Discussion Board, led optional sessions, and held weekly and extra office hours. Developed and administrated a PyTorch-based Image Classifier Machine Learning project assignment.

Student Volunteer - NAACL 2022 May 2022 - Jul 2022

Student Volunteer - ICTD 2022 Jun 2022

Teaching Assistant - CSE473: Introduction to Artificial Intelligence - Jared Moore - University of Washington Spring'22

- Actively engaged in course development, answered questions on Ed Discussion Board, led optional sessions, and held weekly and extra office hours. Designed a new Constraint Satisfaction Problems (CSPs) homework problem with Wordle Game.

PROJECTS

MiniTorch - Python - [Doc] Aug 2023 - Present

- Built MiniTorch, a **Python**-based re-implementation of the Torch API, implementing essential tensor operations, automatic differentiation algorithms, and 1D and 2D convolution and pooling functions.
- Applied parallel programming and GPU acceleration with CUDA for improved performance in deep learning computations, including the development and optimization of matrix multiplication functions.
- Conducted training and implementation of image classifiers using MiniTorch for digit and sentiment classification tasks.
- Engaged in a knowledge-sharing learning environment through real-time feedback and code reviews with peers.

YOYbotInnovation – Python, Git - [Doc][Video] Mar 2023 - June 2023

- Developed and fine-tuned autonomous algorithms for MuSHR cars using ROS in **Python**, incorporating kinematic car motion models, PID/Pure Pursuit/MPC controllers, and Lazy A motion planning.
- Utilized RViz and PlotJuggler for in-depth simulation visualizations and managed documentation via **Git**.
- Collaborated with two teammates for real-time feedback, aligning algorithms with user needs and real-world challenges.
- Conducted extensive real-world testing on MuSHR vehicular platforms, ensuring the integration and seamless performance of localization, planning, and control components.
- Optimized localization, planning algorithms, and controls, enabling the MuSHR car to navigate from start to finish in just 11 seconds, well under the anticipated 30 seconds, with minimal oscillations but no collisions.

Birds? - Python - [GitHub][Kaggle][Slide] Mar 2023 - June 2023

- Fine-tuned the EfficientNetB4 model using **Python** in Google Colab and secured 2nd place out of 32 teams in a class-wide Kaggle competition on bird species identification. Analyzed a 9.95GB dataset and achieved an accuracy of 84%.
- Effectively addressed issues of overfitting and memory constraints.

Byte-Level Deep Pyramid CNN Classifier – Python, PyTorch - [Report][Codebase] Sep 2022 - Dec 2022

- Developed a byte-level text classifier using Squeezed Very Deep Convolutional Networks (SVDCNN) to categorize text inputs with various encodings into English or Corrupted classes. Built model, dataset, and demo using **Python** and **PyTorch**.