Haoyue (Bill) Xiao

+1 (510) 424 1547 - bill.xiao@berkeley.edu

EDUCATION

University of California, Berkeley   **Berkeley, CA, USA**
Computer Science BA & Mathematics, Applied. BA   *August 2021 - May 2025*
GPA: 4.0/4.0
Honors: Phi Beta Kappa (Selected as Junior), Upsilon Phi Epsilon, Dean's List

Core Courses: Multivariable Calculus, Foundations of Data Science, Linear Algebra, Differential Equations, Data Structures, Machine Structures, Algorithms, Machine Learning, Probability Theory, Random Process, Convex Optimization, Real Analysis, Numerical Analysis, Abstract Algebra, Discrete Mathematics **(Straight A/A+'s)**

TECHNICAL SKILLS

Programming Languages: Python, C, Java, JavaScript, SQL, MATLAB, R, LaTeX
Tool/App: PyTorch, Transformer, Git, SciPy, Spacy, Docker, Flask, Azure Cloud, RISC-V
Technique: Data Analysis, Machine Learning, Deep Learning, Reinforcement Learning, Natural Language Processing
OS: Linux, MacOS, Windows, ROS

RESEARCH & COMPETITION EXPERIENCE

SKILLED: Multisensory Human-to-Robot Imitation with Skill Library   **Robotics Institute CMU**
Undergraduate Researcher   *Pittsburgh, PA, USA | May 2024 - Present*
- Equip the Mobile Aloha architecture with audio and tactile sensors.
- Fuse multisensory data into the Action Chunking Transformer (ACT) for more precise goal-conditioned imitation.
- Collect high-quality, goal-conditioned human demonstrations for each skill in the skill library.
- In the wild, interpret first-person one-shot demonstrations into goal-conditioned skills, then execute them in order from the skill library.

Development of an Augmented Chatbot Pipeline   **Supervisor: Prof. Anastassia Fedyk**
Undergraduate Researcher   *Berkeley, CA, USA | Feb 2024 - Present*
- Use Mistral's open-source 7B model to build a RAG-based chatbot pipeline that can generate precise and professional answers based on augmented articles. Build a web app using Docker and deploy it on Azure Cloud.

Anomaly Detection on Medicine Orders   **Supervisor: Prof. Feng Chen (UBC)**
Undergraduate Researcher   *Remote | Dec 2023 - Present*
- Anomaly detection on large-scale real world dataset on the price data of between-corporation medicine purchases. Use regression to predict medicine prices by category and use statistical analysis to detect abnormal orders.
- Use multiple algorithms including multivariable regression, random forest, and autoencoder for cross validation.
- Encapsulate the model and provide user-convenient methods for the client company to use.

Predicting Student Performance from Game Play   **Silver Medal**
Kaggle Competition   *June 2023*
- Predict the player's performance as he plays in an education game with only data from the current session.
- Utilized Pandas for data extraction and leveraged the Polars library to engineer additional features. Create metrics to quantify player engagement and success, including event counts per session, temporal event patterns, and special achievements like "bingo" counts, and the number of times to get special items/reach hidden rooms.
- Use XGBoost model to predict future performance with hyperparameter tuning for highest predict performance.

SELECTED COURSE PROJECTS

- **Design and Implementation of a 2D Tile-based World Generator in Java**, Implemented a world engine that generated completely randomized and unique worlds. Developed a room placement algorithm to ensure non-overlapping random room generation. Employed the Minimum Spanning Tree (MST) algorithm to create hallways connecting the rooms, enabling traversal within the generated world
- **Classification of the MNIST, Spam, CIFAR-10 data**, Implemented various classifiers (e.g. SVM, QDA, LDA, FFNN, and CNN), and fine-tuned them with hyperparameter optimization on the MNIST dataset to achieve high accuracy. Built the ResNet32 model using PyTorch and trained it on the CIFAR-10 dataset. Applied the ResNet32 model to spam detection and incorporated distinguishable features to improve linear separability.
- **Optimizing Matrix Convolution in C**, Implemented a basic matrix convolution algorithm and enhanced its performance by applying various techniques (e.g. SIMD, multithreading, and multiprocessing). Conducted extensive testing and benchmarking to measure the speed and efficiency of the optimized matrix convolution algorithm, resulting in a significant speed improvement of 10x