

Huaiyu Ma

[111-222-3434] | [Blahaaa@gmail.com] | Lanzhou, China

Education

Ohlone College 2027

- Associate of Applied Math

Training

Ohlone College 2025

- Semiconductor Skills Bridge Academy

Skills

- **Git and Github**
- **Programming Language:** C++, Python
- **Data Analytics:** Data collection, recognize structures, interpret data
- **Machine learning algorithm:** Supervised machine learning, unsupervised machine learning, ensembled machine learning, deep learning(CNN, RNN, Transformer)
- **Large Language Model:** LLM architecture, tokenization, embedding, attention mechanism
- **Course Work:** AP Calc BC, AP Statistics, Multi-variable Calculus, Differential Equation, Linear Algebra, Introduction to C++, Introduction to Python

Projects

Facial Recognition Project | *Independent* Oct. 2025

- Developed a Python program using various python libraries to detect and blur individual faces in images for privacy protection.
- Enhanced the program's flexibility by enabling it to process inputs from both static images and live camera feeds.

Image Generation and Stable Diffusion Project | *Independent* Sept. 2025 - Oct. 2025

- Developed a synthetic face generation pipeline using Stable Diffusion v1-5 and Hugging Hub Diffusers library to create labeled emotion datasets.
- Implemented automated prompt engineering with random genders and ethnicity attributes to enhance dataset diversity and realism.
- Optimized model performance and storage by leveraging GPU acceleration(CUDA/MPS), float16 precision, and efficient file management and cleanup for large-scale image generation.

LLM Tokenizer | *Independent* August. 2025 - Oct. 2025

- Create a tokenizer in Python to process a short story, generate token IDs and implement Byte Pair Encoding(BPE) for efficient subwords representation.
- Created a data preprocessing pipeline with sliding windows and Pytorch dataloader for batching and sequencing management during model training.
- Implemented embedding layers, including token and position embedding, to prepare tokenized data for transformer based model input.