Yu-Hui Cao

+64 21 029 47236 | yuhuicao20@gmail.com | LinkedIn | GitHub

Education

University of Auckland
Bachelor of Engineering (Honours)

Experience

Undergraduate Research Assistant - Generative AI & Computer Vision May 2026 – Present
University of Auckland

- Developed an end-to-end Python generative AI pipeline for creating synthetic medical images and annotations.
- Applied synthetic data to improve machine learning performance under limited-data conditions and evaluated results using quantitative metrics.
- Collaborated with researchers to iterate on workflows and improve reproducibility of the data generation process.

Research Intern – Deep Learning & Computer Vision Nov. 2025 – Feb. 2026
University of Auckland

- Developed and evaluated computer vision augmentation strategies for neuronal histology instance segmentation with limited labelled data
- Trained and benchmarked Mask R-CNN models using ResNet-50 and Swin Transformer backbones to quantify augmentation-driven performance improvements
- Explored GAN-based synthetic data generation for histology augmentation and evaluated its impact on downstream segmentation performance

Projects

Research Paper Query Assistant | Python, Llama 4 Scout, Pinecone, FastAPI, React

- Built an end-to-end RAG application for semantic search and question answering over research papers.
- Developed an automated PDF ingestion, embedding and retrieval pipeline using Python and Pinecone..
- Deployed a React and FastAPI interface to make AI-powered document search accessible through a web application.

3D Medical Image Segmentation – Paediatric CT | Python, nnU-Net, DICOM, NIfTI

- Trained a 3D full-resolution nnU-Net for automated paediatric CT bone segmentation, achieving a 0.95 Dice score across 5-fold cross-validation
- Reduced manual preprocessing by automating DICOM-to-NIfTI conversion for 330 medical images.
- Built a voxel-to-mesh reconstruction workflow for generating 3D bone models, achieving a 1.30 mm median surface error

Sentiment Analyser | Python, PyTorch, FastAPI, React.js

- Implemented an encoder-only Transformer from scratch in PyTorch, including multi-head attention, positional embeddings, and residual connections
- Trained the model for sentiment classification and evaluated its performance on user-provided text
- Deployed the model through a FastAPI backend and React interface, enabling real-time sentiment inference

Technical Skills

Languages: Python, JavaScript, SQL, HTML/CSS, R, MATLAB
Machine Learning: PyTorch, scikit-learn, Transformers, HuggingFace, Computer Vision (CV), Deep Learning
Generative AI: Diffusion Models, GANs, Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), LangChain, LangSmith
Web & APIs: FastAPI, React, REST APIs
Data & Tools: Numpy, pandas, Matplotlib, Docker, Git, GitHub
Platforms: Pinecone, Hugging Face, Google Colab, Kaggle, Supabase