

Yi Hao

Data Scientist & Machine Learning Engineer

contact@yihao.au | 0410 455 136 | Sydney, NSW (Australian Citizen)
yihao.au | linkedin.com/in/yi-hao-sydney | github.com/YiHaoFromEarth

Summary

Data Scientist specialising in advanced statistical analysis, machine learning, and deep learning architectures to drive operational efficiency and mitigate risk in complex environments. Combines a robust computer science foundation with rigorous mathematical modelling to engineer scalable, stable pipelines and extract high-value predictive intelligence from messy, large-scale data. A precise technical communicator adept at translating complex algorithmic outputs into actionable business strategy that de-risks executive decision-making.

Skills

- **Data Science & Analytics:** Statistical Analysis & Modelling, Time-Series Forecasting, Mathematical Optimisation, Predictive Modelling, Exploratory Data Analysis (EDA)
- **Machine Learning & Deep Learning:** Neural Network Architectures, Computer Vision, Model Tuning & Hyperparameter Optimisation, Performance Evaluation & Validation
- **Software & Code Craftsmanship:** Object-Oriented Programming (OOP), Modular Software Design, Clean Code Architecture, Version Control (Git)
- **Technical Toolkit & Infrastructure:** Python (PyTorch, NumPy, SciPy, Pandas, Scikit-Learn), C, SQL, Linux, High-Performance Computing (HPC), LaTeX

Professional Experience

Machine Learning Research Engineer (Honours R&D) — *The University of Sydney*

Feb 2026 – Present | Camperdown, NSW

- **System Stability & Risk Mitigation:** Architecting and testing custom deep learning frameworks to systematically identify, predict, and eliminate performance failure points in complex neural networks before deployment.
- **Scalable Infrastructure Engineering:** Developing custom Python and PyTorch simulation workflows to execute large-scale data experiments on high-performance computing (HPC) clusters, optimising background resource allocation and processing efficiency.
- **Advanced Predictive Analytics:** Applying advanced statistical modelling to isolate critical trends and establish high-fidelity reliability thresholds within highly volatile, noisy, and unorganised data environments.

Independent Machine Learning Engineer — *Open-Source Software Development*

Jan 2026 – Present | Sydney, NSW

- **Production-Grade Architecture:** Designed and implemented a modular, object-oriented machine learning framework using standardised data pipelines; built custom, reusable data transformers that automated data preparation and eliminated data leakage risks during validation.
- **Feature Engineering & Data Optimization:** Engineered mathematical rules to combine redundant variables, using structural data scans to compress feature space, eliminate noise, and maximise model processing efficiency.
- **High-Volume Data Aggregation:** Built analytical pipelines to process, clean, and align over 14 million localised time-series data points, ensuring data consistency and accuracy across complex geographical and time-zone shifts.
- **Advanced Noise Filtering & Trend Extraction:** Developed a custom mathematical filtering engine to isolate underlying structural trends and cyclical patterns from highly volatile, noisy background data.

Research Project Fellow (Special Studies Program) — *The University of Sydney*

Aug 2023 – Nov 2024 | Camperdown, NSW

- **Data Integration & Pipeline Engineering:** Built end-to-end Python data pipelines to automate the consolidation of massive, unorganised datasets; implemented advanced statistical frameworks to extract reliable trends and predictive intelligence from incomplete data.
- **High-Cadence Time-Series Analytics:** Engineered automated data-cleaning and filtering systems to process high-frequency time-series datasets; eliminated systemic noise and errors to preserve data integrity and ensure high-fidelity inputs for sensitive predictive models.
- **Automated Computer Vision & Scalable Database Construction:** Designed a production-ready image processing pipeline to automatically identify, classify, and extract features from thousands of raw data candidates, replacing manual processing with scalable, automated database infrastructure.

Education

The University of Sydney | *Sydney, NSW*

Bachelor of Science (Honours) — Expected 2026

- **Specialization:** Theoretical and Computational Physics.
- **Thesis:** Exploring Extended Criticality in Heavy-Tailed Neural Networks.

Bachelor of Science / Bachelor of Advanced Studies — 2025

- **Double Major:** Physics and Computer Science.
- **Academic Performance:** Weighted Average Mark (WAM) of 79.6 / 100.
- **Key Achievements:** Dean's List of Excellence (2023); 3x Special Studies Program (SSP) Research Projects completed at High Distinction.

North Sydney Boys High School | *Sydney, NSW*

Higher School Certificate (HSC) — 2017 – 2022

- **Academic Performance:** Achieved an ATAR of 98.85.

Interests

- **Computational & Data Exploration:** Experimenting with local open-source models, fine-tuning large language models on personal infrastructure, and discovering patterns in complex public datasets.
- **Hardware Architecture:** Designing and assembling high-performance PC systems optimised for heavy parallel processing and machine learning workloads.
- **Observational Astronomy & Photography:** Practicing amateur astronomy using Dobsonian telescopes and capturing astronomical events through digital SLR photography.