

Hou Hin Ip (Izaak Ip)

Email: izaacip816@gmail.com Website: www.izaacip.com
linkedin.com/in/houhinip/ | github.com/Picaazi

PROFILE

MSc Computing (AI & ML) student at Imperial College London with experience developing **AI agents**, **computer vision models**, and **ML evaluation** and **data pipelines**. Skilled in **model development**, **performance optimisation**, and building reliable AI systems across research and industry.

EDUCATION

Imperial College London **London, UK**
MSc Computing (AI & Machine Learning) **Sep 2026 - Sep 2027**
Relevant Modules: Computer Vision, Generative AI, NLP, Reinforcement Learning, World Model
Research projects: ISO project; individual MSc research project (planned).

University of Bristol **Bristol, UK**
BSc Data Science (with First Class Honours) **Sep 2023 - Jun 2026**
Relevant Modules: Information Theory, Statistical Machine Learning, Time Series, Bayesian Modelling
Bachelor's research project: Diffusion models as plug-and-play priors for image restoration.
Received offer for **Cambridge MPhil Data Intensive Science**

TECHNICAL SKILLS

Programming Language: Python, C++, SQL, R, C, JavaScript
Frameworks & Libraries: PyTorch, Scikit-learn, NumPy, Pandas, Hugging Face Transformers
Machine Learning: Computer vision, LLMs, RAG, multimodal learning, diffusion models, NLP
ML Engineering: Model training and evaluation, experiment tracking, HPC
Data Science & Processing: Data cleaning, exploratory data analysis, feature engineering
Software & Data Tools: Git, AWS, Google Cloud, REST APIs, ETL pipelines, MCP
Languages: English (Fluent), Mandarin (Native), Cantonese (Native)

PROFESSIONAL EXPERIENCE

NatWest **London, UK**
Engineering Summer Intern **Jun 2026 - Aug 2026**

- Led development of a **technology recommendation AI agent** using **Microsoft Power Platform**.
- Developed **automated evaluation scripts** to analysis **AI-agent conversations**, identify agent's error, and access accuracy with **reference data**, summarising into a **evaluation matrix**.
- Developed **automation scripts** using **Kiro** and **Model Context Protocol (MCP)** to streamline development and testing.
- Collaborated in an **Agile Scrum** team using **Jira** and **Confluence**.

BGI Genomics (STOmics Tech) **Shenzhen, China**
Algorithm Test Engineer Intern **Jul 2025 - Sep 2025**

- Managed **data cleaning and annotation pipelines** for large-scale biomedical imaging datasets used to train **YOLO computer vision models**.
- Optimised **Python data-quality validation pipelines**, reducing processing time from **2 minutes to 15 seconds per image (8x speedup)**.
- Led and **standardised annotation workflow** to ensure dataset quality, producing **2,000+ samples**.
- Built **QC dashboards** with **OpenCV**, **Pyvips**, **Pillow** and **Openpyxl** to visualise defect maps, reducing review time from **5 minutes to under 2 minutes per image**.

HeadStart Global (Startup Lomara) **Bristol, UK**
Associate Consultant **Feb 2026 - Apr 2026**

- Developed **YOLOv8 model-training pipelines** for a smart-bin **computer vision detection system**, improving accuracy by **20%** compared with the baseline.

180 Degrees Consulting Bristol **Bristol, UK**
Data Science Client Service Lead **Sep 2024 - May 2026**

- Led **54 student consultants** and **12 data science consulting projects** for NGOs and startups.
- Deployed a **full stack data system** using **React**, **Flask** and **MySQL** on **Google Cloud** for client data.
- Developed an **interactive dashboard** and **scalable data pipeline** to support client reporting. ([GitHub](#)).

Citic Wealth Management
Data Science Intern

Shanghai, China
Jul 2024 - Aug 2024

- Built **XGBoost forecasting model** to predict product demand and customer preferences.
- Used **Python** and **SQL** to analyse financial sales and customer data and support portfolio decision making.

HeadStart Global (Startup OmniMesh)
Associate Consultant

London, UK
Sep 2026 - Oct 2026

- Developed a website strategy to present OmniMesh products to commercial and technical users.

AI & MACHINE LEARNING PROJECTS

Bias Mitigation in Retrieval Augmented Generation ([GitHub Link](#)) **Jun 2025 - Sep 2025**

- Developed a **RAG pipeline** with **PyTorch**, **Sentence Transformers** embeddings, **dense** and **BM25 retrieval**, and LLM-based query rewriting and decomposition.
- Implemented **evaluation pipeline** comparing query strategies against original-query retrieval with fixed random seeds and top-k settings; measured semantic Recall@k, **MRR**, document overlap and runtime.
- Tracked experiment configurations and outputs with **Weights & Biases** and **CSV**, and implemented demographic representation and bias-amplification metrics to examine retrieval behaviour.

KidSat Extension , Summer Research Bursary Project ([arxiv](#)) **May 2025 - Jul 2025**

- Contributed to **multimodal model training** combining **DINOv2** vision transformer embeddings with **spherical-harmonic geographic encoders** for satellite-based childhood poverty estimation.
- Ran **HPC** training and evaluation workflows, comparing image-only and geographic features, regression heads (**ridge**, **MLP**, **XGBoost**, **LightGBM**) and training settings using **five-fold cross-validation**.
- Supported **image-quality screening** and dataset cleaning; the new model reduced **MAE** from **0.2167** to **0.1759** (**18.83% improvement**) on the original dataset and achieved **0.1658** across **33 countries**.

Diffusion Models for Image Inverse Problems ([PDF](#)) **Sep 2025 - Mar 2026**

- Integrated pretrained **score-based diffusion models** with **ADMM** and **Split Gibbs Sampling (SGS)** for **plug-and-play image restoration**, combining generative priors with data-fidelity constraints.
- Built an evaluation framework for Gaussian and motion deblurring and **4× super-resolution** on **50 held-out FFHQ images**, using **PSNR**, **SSIM** and **LPIPS** to compare reconstruction quality.
- Compared splitting-parameter schedules and stochastic, deterministic and hybrid **ADMM** updates; **SGS outperformed ADMM** across all three simulated image-restoration tasks.

ADDITIONAL ACADEMIC EXPERIENCE

Subject Editor, Student Research Journal (two editions from 2024-2026)

Course Representative, Data Science BSc in University of Bristol (2025-2026)

Teaching Support: Algorithms and Machine Learning; Advanced Linear Modelling and Classification (2025-2026)

HOBBIES

Game development, computer games, reading, board games, ACG and gym training.