

# Gavin Yue

📍 London 📩 zhilinggavin@gmail.com ☎ +44 7397682706 💬 LinkedIn 📚 Google Scholar

## PROFESSIONAL SUMMARY

Machine Learning Researcher advancing **cutting-edge Generative AI (GenAI) research** while **engineering end-to-end ML model development**, covering data preprocessing, model development, benchmarking, and deployment. Strong expertise in **Computer Vision, NLP, and large-scale AI architectures** (GANs, Diffusion Models, Transformers, LLMs), with hands-on experience in **MLOps and cloud deployment** (Docker, MLflow, AWS S3, SageMaker). Published **2 peer-reviewed AI papers** and developed **AI solutions for healthcare**. Passionate about **bridging research and real-world applications**, optimising AI models for efficiency, scalability, and production deployment.

## EDUCATION

### Imperial College London, UK

Sept 2023 – Oct 2024

*MRes in Machine Learning for Medical Applications (**Distinction**)*

- Research Focus: Generative AI, Computer Vision, Data Processing
- Achievements: Published **2 peer-reviewed Papers**, invited presentation at *Hamlyn Symposium*

### University of Birmingham, UK

Sept 2021 – July 2023

*BEng in Mechanical Engineering (1st Class Honours)*

- Thesis: **Cloud AI solutions** for industry robots using **AWS and LLMs**
- Key Skills: Engineering design, **cross-functional collaboration** (30+ members), CAD, FEA

### Wuhan University of Technology, China

Sept 2018 – July 2021

*BEng in Mechanical Engineering (1st Class Equivalent)*

- **Awards:** Excellent Student Leader award; Academic Excellence Scholarship; Chair of Class Committee

## TECHNICAL SKILLS

**Programming:** Python, C/C++, MATLAB, MySQL, Linux (HPC, Shell)

**AI/ML:** PyTorch, TensorFlow, Scikit-Learn, CNNs, LLMs, GANs, Diffusion, Transformers

**Cloud & MLOps:** Docker, Git, W&B, AWS (S3, SageMaker, lambda), MLflow

**Data Engineering:** Pandas, NumPy, OpenCV, SQL, large-scale data processing, pipeline automation.

**ML development:** End-to-end ML model lifecycle, CI/CD, cloud computing

## AI RESEARCH EXPERIENCE

### Machine Learning Researcher, Imperial College London

Sept 2023 - Present

*Skills: GANs, LLM, Computer Vision, AI Model Benchmarking*

- **Developed and deployed** AI-powered healthcare solutions, surpassing commercial benchmarks (Meta SAM) in accuracy by 40%
- **Optimised large-scale AI models** (GANs, Diffusion Models, Vision Transformers, and LLMs) for real-world applications, ensuring efficiency across multiple evaluation metrics
- **Integrated NLP & Vision** models into a **production-ready AI tool** for disease detection and clinical documentation automation
- **Benchmarked and fine-tuned** ML models, improving inference speed and reducing model latency for real-time AI applications
- Built scalable **ML pipelines** integrating Docker, MLflow, AWS S3, and SageMaker, optimising model deployment efficiency by 30%
- Created and processed a **200,000+ image dataset**, implementing data warehousing, preprocessing, and augmentation to improve model accuracy by 35%
- **Published 2 peer-reviewed AI Papers** in IJCAI 2025 and IEEE ISBI 2025, presenting findings to a global audience of 200+ researchers at the Hamlyn Symposium

**Key Research Projects:**

- *GenAI for Semantic Segmentation*: Developed and deployed an **AI auto-segmentation tool** to assist radiologists, enhancing workflow automation in clinical imaging
- *Multimodal NLP for Clinical Reports*: Developed a **Vision-Language AI system** (vision + text) for automated clinical documentation and diagnostics

#### AI & Robotics Research Project, University of Birmingham

*July 2022 - Sept 2023*

*Skills: AWS, LLM, Cloud, IoT*

- Developed an **LLM-powered robotic assistant** for industrial automation, applying **Reinforcement Learning** for adaptive decision-making
- Engineered a **cloud-based AI framework** (AWS Lambda, AWS IoT Core) for real-time robotic deployment, reducing latency by 40% through optimised speech recognition

### INDUSTRY & INTERNSHIP EXPERIENCE

---

#### Machine Learning Engineer Intern, AI Consulting Startup, UK

*Feb 2025 – Present*

- Developing an automated pipeline for **Company Insight Reports**, leveraging LLMs for **business growth and risk evaluation**
- Extracting and processing financial and operational **public data** using web scraping, **Named Entity Recognition (NER)**, structuring information in Excel and databases for scalable analysis
- Applying **LLaMA2** for **automated trend analysis**, sentiment evaluation, and risk assessment, providing data-driven industry insights

#### Machine Learning & Robotics Intern, Imperial College London

*June 2024 - Sept 2024*

- Developed real-time **SLAM** (ORB-SLAM2, EKF) for marine robot **localisation and mapping**
- Integrated **Stereo vision and CNNs** (StereoNet) for 3D underwater scene reconstruction
- Built dual simulations with **ROS** and Gazebo, training **reinforcement learning** (DDPG) and integrating MPC with path planning for optimised trajectory planning and obstacle avoidance

#### Machine Learning Intern, Bright Network, UK

*July 2022 - July 2022*

- Developed and deployed **AI models in AWS** during Amazon-led ML workshops.
- Designed Reinforcement Learning-based navigation algorithms for autonomous systems

### PUBLICATIONS

---

#### Enhancing Weakly Supervised Semantic Segmentation for Fibrosis via Controllable Image Generation ↗

2024

*IEEE International Symposium on Biomedical Imaging (ISBI 2025)*

*Zhilong Yue, Yingying Fang, et al.*

#### Decoding Report Generators: A Cyclic Vision-Language Adapter for Counterfactual Explanations ↗

2024

*International Joint Conference on Artificial Intelligence (IJCAI 2025)*

*Yingying Fang, Zihao Jin, Zhilong Yue, et al.*