

Pugalenthi Magendran

FORWARD DEPLOYED AI ENGINEER | APPLIED AI & ML SYSTEMS

Melbourne, VIC • +61 431 235 427 • pugalenthi0928@gmail.com • [LinkedIn](#) • [GitHub](#)

Turning complex business requirements into secure, scalable and production-ready AI systems.

TECHNICAL SKILLS

Programming Languages

Python, SQL, Bash

ML and Deep Learning

PyTorch, scikit-learn, NumPy, Pandas, CNNs, ResNet-style encoders, Transformers, vision-language models, computer vision pipelines

LLMs and Generative AI

LLaMA, Qwen 2.5, FLAIR, OpenCLIP, BiomedCLIP, OpenAI APIs, LoRA and PEFT fine-tuning, embeddings and vector stores, prompt engineering, LLM-as-judge and ROUGE evaluation

Engineering and MLOps

FastAPI, JWT, bcrypt, Pydantic, SQLAlchemy, Streamlit, Docker, LangGraph, PostgreSQL, JSONL pipelines, Git, CI/CD, Jupyter, Colab, AWS, Azure, MLX and Apple Silicon

EDUCATION

Master of Artificial Intelligence (CGPA

3.7/4.0): Monash University, Dec 2025

- **Thesis:** “From Prompts to Probes: Zero-Shot and Few-Shot Transfer of Foundation Models for Retinal Fundus Classification,” supervised by Dr Yasmeen George.

Bachelor of Business (Finance) – Minor:

Accounting: Monash University, May 2023

CORE STRENGTHS

- AI Engineering & Solution Development
- Large Language Models (LLMs) & Generative AI
- Foundation Models & Transfer Learning
- AI Workflow Design & Automation
- Machine Learning & Deep Learning
- Computer Vision & Medical AI
- Backend Development & API Engineering
- AI Security, Privacy & Governance
- Prompt Engineering & Evaluation
- Training Data Generation & Fine-Tuning
- Startup Intelligence & Commercial Analysis
- Analytical & Critical Thinker

PROFESSIONAL SUMMARY

Highly analytical and technically proficient professional with a Master of Artificial Intelligence and hands-on experience designing, developing, and evaluating foundation models, large language model applications, AI workflow automation systems, and secure AI-enabled solutions. Strong technical capability across machine learning, generative AI, computer vision, backend engineering, API integrations, data pipelines, and MLOps, complemented by commercial experience spanning client delivery, venture capital, startup ecosystems, and digital transformation. Able to translate complex business and technical requirements into practical end-to-end AI solutions, from privacy-aware LLM applications and authentication services to production-oriented AI workflows, training-data pipelines, model evaluation frameworks, and scalable system architectures.

SELECTED TECHNICAL ACHIEVEMENTS

- **Selected into Perplexity AI’s inaugural Global Fellowship**, collaborating with a global cohort spanning 30+ countries to explore enterprise AI adoption, AI-assisted research workflows, and commercial applications of generative AI.
- **Designed and executed more than 63 controlled foundation-model experiments** across retinal disease classification benchmarks, evaluating zero-shot prompting and few-shot adaptation approaches using approximately 3,000 labelled ophthalmology images.
- **Achieved up to 0.921 AUROC for glaucoma detection and 0.735 Macro-F1** for diabetic retinopathy classification, outperforming OpenCLIP and BiomedCLIP zero-shot baselines under a controlled retinal foundation-model benchmark protocol.
- **Created a privacy-preserving LLM middleware layer** that detected and anonymised sensitive information, including Tax File Numbers, Medicare numbers, passport numbers, student identifiers, financial data, and contact information before model submission.
- **Developed a secure AI gateway authentication service** using Python, FastAPI, JWT authentication, SQLAlchemy, bcrypt, and audit logging, protecting access to internal AI applications, conversation histories, model endpoints, and administrative functions.
- **Built an LLM training-data factory that transformed source documents into 200 high-quality JSONL training examples** across question answering, summarisation, instruction-following, and reasoning tasks.
- **Fine-tuned Qwen 2.5 using a custom-generated training dataset**, reducing validation loss from 2.345 to 0.730 and improving ROUGE-L performance from 0.289 to 0.417 following 120 training iterations.

CAREER EXPERIENCE

ZeroFold — Melbourne, Australia

Jan 2026 – Present

Forward Deployed AI Engineer (Contract)

- Led technical discovery and **solution architecture** across engagements involving **NAB, Owen Hodge Lawyers, Alteris Financial Group, Costi Cohen, DexIQ and Marine Biomedical**, translating complex stakeholder requirements into production architectures, API specifications, implementation roadmaps and acceptance criteria.
- Architected an **agentic email system** using **LangGraph, Microsoft Graph, Pipedrive, PostgreSQL**, queue workers and human review to automate triage, CRM updates and response control.
- Build **RAG applications, AI workflows and secure FastAPI** services using Python, LLM APIs, vector retrieval and Docker, with privacy, audit and evaluation controls.
- Direct engineers through **implementation, testing, deployment and handover**, balancing model quality, latency, security, reliability, infrastructure cost and maintainability.

Perplexity AI – Remote (San Francisco, USA)

March 2025 – Jun 2026

Business Fellow & Campus Ambassador, AI Product & Research

- Designed and tested AI workflow prototypes using Perplexity Enterprise Pro, focusing on company intelligence, research synthesis and content operations.
- Built and evaluated AI-assisted research systems across finance, healthcare, legal, technology, venture capital, and content operations use cases.
- Assessed trade-offs between single-prompt and staged reasoning workflows, identifying improvements in source coverage, output structure, consistency, and reusability.
- Created repeatable prompt frameworks and workflow templates that improve consistency across research, content generation and intelligence gathering activities.
- Participated in masterclasses and workshops with global technology leaders, including NVIDIA CEO Jensen Huang and Box CEO Aaron Levie to strengthen understanding of enterprise AI deployment and adoption strategies.

Monash University, Melbourne VIC

Feb 2025 – Nov 2025

AI Researcher, Foundation Models for Ophthalmology

- Designed a controlled benchmarking framework comparing zero-shot prompting and few-shot linear probing approaches across diabetic retinopathy, glaucoma, cataract, and pathological myopia classification tasks.
- Engineered reproducible Python and PyTorch experimentation pipelines incorporating retinal image preprocessing, quality filtering, patient-level dataset partitioning, cached embedding generation, and configuration-driven experiment execution.
- Evaluated foundation model behaviour across MESSIDOR, REFUGE, and ODIR datasets using varying supervision budgets (5%, 10%, and 20% labelled data) to assess adaptation efficiency in data-constrained environments.
- Implemented comparative benchmarking workflows for FLAIR, OpenCLIP, BiomedCLIP, and ResNet-based encoders to analyse representation quality and downstream classification performance.
- Investigated the impact of prompt design, quality, feature separability, and label availability on retinal disease classification outcomes.
- Developed experiment-tracking and reproducibility mechanisms enabling end-to-end regeneration of model outputs, evaluation metrics, and benchmark comparisons from a single configuration workflow.
- Produced evidence supporting the viability of foundation-model transfer learning for ophthalmology applications where labelled clinical data is limited and expensive to obtain.

AUG Melbourne, Melbourne VIC

2024 – 2025

Digital Systems and Applied AI Developer (Internship)

- Contributed to internal digital transformation initiatives by developing backend services, AI governance capabilities, and workflow automation solutions supporting staff operations.
- Engineered a secure authentication and access management layer for an internal AI gateway, enabling controlled access to conversational AI services, model selection, user management, and audit functionality.
- Designed a privacy-first middleware that screened prompts for sensitive Australian and student-related information before submission to public LLMs, supporting safer AI adoption within the organisation.
- Implemented prompt anonymisation and data restoration workflows to balance privacy requirements with usability.
- Collaborated with business stakeholders to translate operational requirements into practical software, automation, and applied AI solutions.

1337 Ventures, KL

2022

Venture Capital Analyst

- Mapped and screened approximately 400 Southeast Asian startups across 8+ VC screening dimensions, including sector, geography, business model, founder background, funding stage, traction signals, competitive positioning, and investor activity, building roughly 3,000+ structured sourcing data points.
- Maintained the Start-up List and supplied market maps, startup shortlists, and company intelligence used by the investment team to identify high-potential companies and support deal flow for the accelerator and investment pipeline.
- Conducted investment due diligence across fintech, biotech, SaaS and other sectors, evaluating market opportunity, competitive positioning, business models, traction, founder quality and key commercial risks to support investment decisions.

AI ENGINEERING PORTFOLIO

Training Data Robo: LLM Training-Data Factory and Dashboard

2025

Python, PyTorch, MLX, Streamlit, Docker, LoRA, Qwen 2.5, LLM-as-judge evaluation

- Built a configurable training-data generation platform that processed six source documents into 200 JSONL training examples across question answering, summarisation, instruction-following, and reasoning tasks.
- Developed automated quality-scoring, deduplication, difficulty-calibration, train/test splitting, and LLM-as-judge evaluation workflows.
- Generated a 160/40 train-test split with an average LLM-as-judge quality score of 4.77/5.0.
- Used the generated dataset to fine-tune Qwen 2.5 0.5B using LoRA, reducing validation loss from 2.345 to 0.730 after 120 training iterations.
- Improved ROUGE-1 from 0.418 to 0.586 and ROUGE-L from 0.289 to 0.417 through parameter-efficient fine-tuning.