Muhammad Haziq Ilham Aziz Azhar

haziqilhamwork@gmail.com | +60 11 - 31512579 | LinkedIn | GitHub | Kuala Lumpur, Malaysia

SUMMARY

Junior Full Stack Software Engineer with production experience building VLM inference pipelines, object detection systems, and structured evaluation frameworks for real-world document understanding tasks. Proficient in end-to-end ML development, from dataset construction and model benchmarking to deployment and system optimization. Currently contributing to a commercial AI product at Rosary Labs, with a strong research foundation in reinforcement learning and deep learning.

SKILLS

Programming Languages: Python, R, C++, SQL, TypeScript, JavaScript (ES6+), Bash
Machine Learning & Deep Learning: PyTorch, TensorFlow, HuggingFace Transformers, Vision-Language Models (VLMs), YOLO, RF-DETR, CNNs, LSTMs, Supervised & Unsupervised Learning, Time-Series Forecasting, Reinforcement Learning (DQN, PPO, PPO-LSTM), Transfer Learning, LoRA / PEFT
NLP & LLM Tooling: RAG, OpenAI API, Anthropic Claude API, LangChain, AutoGen, Ollama, Embeddings, Text Classification
MLOps & Deployment: Docker, Git/GitHub, FastAPI, Flask, CI/CD Pipelines, Model Serving, RunPod, pytest, Sentry, OpenTelemetry
Cloud & Infrastructure: AWS (S3, EC2), Azure (ACR), Redis, Supabase, MongoDB, SQLite
Data Engineering: ETL Pipelines, Pandas, NumPy, Polars, Data Preprocessing
Front-End: React.js, Next.js, Tailwind CSS, Streamlit, Plotly Dash, WebSocket, REST API
Languages: English (Fluent), Malay (Fluent)

WORK EXPERIENCE

Junior Full Stack Software Engineer, Rosary Labs **January 2026 - Present**

- Architected a structured evaluation framework decomposing model performance into **Retrieval, Extraction, and Computation** metrics for systematic error analysis in quantity audit tasks.
- Refactored AI inference pipeline from parallel to chained **Object Detection-to-VLM** architecture, ensuring deterministic bounding-box-to-text alignment with structured persistence via Django ORM.
- Benchmarked **Table Transformer (TATR), CascadeTabNet, and TableNet**; achieved **82.1s** full-document detection on 92-page inputs with TATR and recommended it based on performance–cost tradeoff over VLM-based extraction.
- Designed an AI-assisted **BOQ population workflow** with empty-field detection, fuzzy description matching, cross-sheet value extraction, and automated aggregation; evaluated **Claude Code SDK vs. Gemini** for out-of-Excel computation.
- Consolidated VLM calls into **single batched API requests** per grid image, reducing overhead while preserving coordinate-tag associations for downstream auditability.
- Prototyped segmentation experiments with **Meta SAM** via HuggingFace Transformers, exploring prompt-based detection and embedding reuse for efficient multi-prompt inference.
- Built a **PDF-to-Excel extraction system** for 92-page engineering drawings: page-to-image conversion, interactive bounding box UI, Gemini-powered table extraction, and automated multi-sheet Excel consolidation via Pandas/Polars.
- Migrated model deployment from Docker Hub to **Azure Container Registry (ACR)** with secure authentication for RunPod, eliminating external exposure risks in production.
- Led **TTL UI revamp** enabling user-assisted table-to-BOQ mapping with dynamic tab separation, drawing-to-table synchronization, and Excel-like formula traceability.

- Refactored legacy Django frontend into **modular component-based architecture**, separating JS/CSS, de-duplicating components, and improving long-term maintainability.

Data Science Intern, Rosary Labs **September 2025 - December 2025**

- Rewrote the core pattern-matching engine, achieving a **160× speedup** in hyperparameter tuning through caching optimization, Pandas-to-NumPy migration, and large-scale synthetic event pair generation across minute-level datasets spanning 2022–2025.
- Designed and executed a full **Object Detection + VLM pipeline** for extracting engineering tags from large-scale P&ID diagrams, integrating YOLO / RF-DETR as a preprocessing stage for robust text localization.
- Built experimental frameworks comparing VLM-only vs. detection-assisted VLM extraction, demonstrating measurable performance gains and identifying failure modes through systematic gap analysis.
- Led dataset construction and annotation for large A1-scale engineering drawings; evaluated detection performance using **mAP@0.5, precision, and recall**.
- Designed a scalable ML inference architecture proposing RF-DETR as a standalone **RunPod serverless** service, enabling efficient batching, reduced memory footprint, and faster startup via lazy-loading.
- Implemented advanced post-processing: grid-based image segmentation, bounding-box cropping, rotation correction for vertical text, and batch inference optimization.
- Diagnosed and resolved **Redis OOM bottlenecks**, profiling memory-intensive functions to improve backend reliability and throughput.
- Utilized **STUMPY** for time-series similarity search and anomaly detection, improving stability of correlation-based sensor monitoring.

PROJECT

Cognitive Temporal Reinforcement Learning with Surprise-Based Learning Rate Modulation

- Developed a biologically-inspired reinforcement learning system that adapts learning rates based on environmental surprise, drawing from neuroscience theories of temporal perception.
- Implemented a forward prediction model using PyTorch to compute prediction errors, integrated Pearce-Hall associability theory for temporal smoothing, and modulated PPO (Proximal Policy Optimization) learning rates via Stable-Baselines3 on Gymnasium's LunarLander-v3 environment, achieving a 14.8% performance improvement over baseline with reduced learning variance.

Curiosity-Driven Exploration Framework for Sparse Reward Environments in Reinforcement Learning

- Developed a comprehensive reinforcement learning research framework implementing 20+ state-of-the-art intrinsic motivation methods for solving hard exploration problems in sparse reward environments.
- Implemented prediction-based curiosity (RND, ICM, DRND from ICML 2024, NGU, BYOL-Explore), count-based methods (SimHash, Go-Explore), skill discovery algorithms (DIAYN, DADS), multi-agent curiosity systems, and LLM-guided exploration using PyTorch and Stable-Baselines3, with custom Gymnasium environments (DeceptiveMaze, KeyDoor, MontezumaLite) for benchmarking exploration efficiency.

Smart Street Object Detection Using Artificial Intelligence of Things (AIOT)

- Developed an AI-powered system to detect and classify street objects in real time using computer vision and IoT technologies.
- Utilized Roboflow for dataset preparation, YOLOv8 for object detection, and deployed the model on ESP32-CAM using Google Colab for training and optimization, enhancing road safety through smart urban monitoring.

Optimizing User Allocation to the Tower Using Cuckoo Search Algorithm

- Developed a metaheuristic-based solution to optimize user-to-tower allocation in cellular networks by focusing on two key performance metrics, signal strength and user proximity to towers.
- Leveraged the Cuckoo Search Algorithm to maximize signal reliability and minimize user-to-tower distance (within a range of 1 to 10 towers), aiming to enhance network efficiency, reduce signal attenuation, and improve overall service quality and user satisfaction.

EDUCATION

Universiti Teknologi MARA (UiTM) **Shah Alam, Selangor**

Bachelor of Information Systems (Hons.) Intelligent System Engineering October 2022 - February 2026

- Cumulative GPA: 3.64
- Dean's List recipient for 6 consecutive semesters in recognition of academic excellence.
- Head of Entrepreneurial Exco for the Artificial Intelligence Society (AIS) UiTM.
- Final Year Project : Smart Street Object Detection using Artificial Intelligence of Things (AIOT)

ACTIVITIES

Artificial Intelligence Society (AIS)

Head of Entrepreneurship

- ProSolve National 2025 – Served as Head of Sponsorship, leading a team that successfully secured up to RM11,000 in sponsorship for a national-level competitive programming event involving universities across Malaysia.
- AI Tech Talk & Learning Revolution at UiTM – Acted as Vice Program Director for a campus-based industry talk in collaboration with ASUS Malaysia, bridging academic interests with real-world applications in AI and innovation.

CERTIFICATIONS AND QUALIFICATIONS

- Microsoft Certified: Azure Data Fundamentals
- Asia Pacific University: Data Analysis with Power BI

REFERENCES

Available upon request.