

Nisyhaal Tamil

AI & Robotics Engineer



✉ nisyhaaltamil6@gmail.com ☎ 601116301017
🌐 [linkedin.com/in/nisyhaal-tamil/](https://www.linkedin.com/in/nisyhaal-tamil/) 🔗 nisyhaal.github.io
📄 github.com/nisyhaal

Profile

AI and robotics engineer with experience in designing and deploying intelligent systems across the full engineering lifecycle from model development and optimization to production infrastructure and customer-facing delivery.

Proven track record of building end-to-end solutions spanning edge AI, computer vision, real-time robotics, and generative AI, with deep expertise in bridging research-level techniques to deployment-ready products on resource-constrained hardware.

Driven by a dual ambition: to push the technical frontier in AI and autonomous systems, and to grow into a leadership role that shapes engineering strategy, mentors talent, and steers high-impact product decisions. Seeking opportunities at the intersection of deep technical ownership and team leadership.

Professional Experience

Software Optimization & Enabling Engineer, Intel 06/2024 – Present | Penang, Malaysia
Designed, developed, and deployed edge AI solutions spanning computer vision, robotics, generative AI, and model optimization on Intel hardware (Arc GPUs, Xeon, NPUs). Work covers the full engineering lifecycle — from training and optimizing ML models to building production-grade backend services, web frontends, and containerized deployment infrastructure.

Edge AI & Model Optimization

- Deployed multi-stage inference pipelines spanning object detection, OCR, speech, and vision-language models on Intel Arc GPUs, integrated GPUs, and NPUs using OpenVINO.
- Quantized and optimized models (INT4/INT8/FP16) and accelerated LLM fine-tuning on Intel XPU for Llama, Qwen, Mistral, and DeepSeek families.

Robotics & Real-Time Systems

- Developed ROS 2 packages for industrial robotic arms, hardware drivers, teleoperation, gripper control, camera integration, and motion planning.
- Built simulation environments for synthetic data generation and sim-to-real transfer; ported modules of a 138K-line C++ control framework from ROS 1 to ROS 2.

Generative AI & Agentic Systems


- Built LLM-driven agentic interfaces with tool-calling and RAG for manufacturing automation and technical document reasoning.
- Developed multi-modal retrieval pipelines and real-time multi-language speech translation on Intel edge GPUs.
- Contributed full-stack features and automated installer tooling to an enterprise LLM fine-tuning platform.

Full-Stack & Infrastructure

- Built full-stack applications (FastAPI, Next.js, React, Gradio) with PostgreSQL, MongoDB, and SQLite backends.
- Designed microservice architectures (up to 10-service Docker Compose stacks) with task queues, reverse proxies, TLS, JWT/RBAC auth, and CI/CD pipelines.

Customer Engagement

- Delivered live technical demos of edge AI and robotics solutions to Intel customers and partners at industry events including Embedded World 2025.

AI Engineer, *Innorative* 

04/2023 – 05/2024 | Kuala Lumpur, Malaysia

Experience Highlights

- Led development of an automated AI-based image classification system for parcel approval and rejection.
- Built a modular text classification system with a FastAPI application for efficient processing.
- Developed a resume parser for PDF and image formats using OpenAI GPT API.
- Contributed to internal tools for image classification model training.

AI Strategy & Business Alignment

- Aligned AI initiatives with business goals to drive operational impact and ROI.
- Assessed business needs and proposed data-driven AI solutions in collaboration with stakeholders.

Pipeline & Infrastructure

- Optimized and managed AI pipelines for efficient data flow and model training.
- Partnered with DevOps to maintain infrastructure and ensure system reliability.

API & Application Development

- Designed and integrated scalable APIs for cross-system functionality.
- Developed and deployed internal applications with focus on scalability and performance.

Machine Learning Engineer, *Glueck Technologies*

06/2022 – 03/2023 | Petaling Jaya, Malaysia

Experience Highlights

- Designed and deployed an end-to-end computer vision pipeline for object detection, OCR extraction, and damage identification.
- Built and deployed scalable Vision AI applications across edge, on-premises, and cloud environments.
- Developed high-accuracy, production-ready AI models using NVIDIA TAO Toolkit.
- Optimized AI deployment using NVIDIA DeepStream and TensorRT.

AI Development & Engineering

- Developed Python-based pipelines for data preprocessing, model training, and workflow automation.
- Built high-performance applications using C++, OpenCV, and video processing techniques.
- Leveraged CUDA and cuDNN for accelerated model performance and efficient computation.

Multi-Platform Deployment

- Delivered AI solutions across edge devices, on-prem infrastructure, and cloud platforms.
- Ensured scalability and performance across diverse deployment environments.

Quality Assurance & Optimization

- Conducted rigorous testing for edge cases, usability, and system reliability.
- Continuously improved model performance and system efficiency through iterative optimization.

Junior Artificial Intelligence Engineer, *SkyMind*

06/2021 – 05/2022 | Bangsar

Experience Highlights

- Developed and enhanced computer vision algorithms for custom use cases, including segmentation, OCR, and fraud detection.
- Designed and deployed deep learning models for face recognition, object segmentation, and document processing.
- Collaborated with business teams to deliver AI-driven solutions aligned with product and client needs.
- Provided technical support and translated model outputs into actionable insights for stakeholders.

AI Research & Development

- Built and optimized deep learning models from prototype to production-ready systems.
- Improved existing algorithms and developed new approaches for complex vision tasks.

Collaboration & Business Alignment

- Worked closely with cross-functional teams to maximize product value through AI solutions.
- Bridged technical implementation with business requirements and client expectations.


Testing & Documentation

- Conducted feature and functional testing to ensure system reliability and performance.
- Produced clear technical documentation to support development and knowledge sharing.

Training & Knowledge Sharing

- Delivered training sessions on AI concepts and system usage for internal teams and external stakeholders.

Education

Master of Data Science, Universiti Malaya 
Grade: 3.95

10/2023 – 06/2024 | Kuala Lumpur, Malaysia

Skills Highlights:

- Data Analysis: Proficient in uncovering patterns within complex datasets.
- Machine Learning: Experienced in developing models for classification, regression, and forecasting.
- Deep Learning: Skilled in implementing LSTM networks for temporal data analysis.
- Distributed Computing: Knowledgeable in building and optimizing cloud-based data processing platforms.
- Programming: Proficient in Python for data manipulation and analysis, with additional experience in R programming.
- Project Management: Experienced in leading project teams and delivering solutions on schedule.

Projects:

- Cerebral Stroke Prediction: Built classification models to predict stroke risk factors.
- Temporal StormGuard: Developed a real-time weather alert system using deep learning.
- Predicting Public Transport Ridership: Analyzed factors impacting ridership patterns in Kuala Lumpur.
- Image Processing on AWS: Led the development of a distributed image processing platform for efficient analysis on AWS Cloud.

Bachelor of Engineering (Honours) (Mechatronics Engineering), Universiti Sains Malaysia

2015 – 2020 | Nibong Tebal, Malaysia

Possess clear understanding of Mechatronic Systems & Design, Robotics & Intelligent Systems (AI) and Control System Design

Dean List - Semester 1 Academic Year 2019/2020

Final Year Project: Flex Force Smart Glove

Individual work of developing a smart data glove to measure the human finger flexion, orientation in space and display them in a smartphone application (Blynk) using Internet of Things (IoT).

Projects

Intelligent Robotic Pick-and-Place System, Sole Author (51/51 commits)

- Architected **end-to-end robotic pick-and-place system** in ROS 2 Jazzy with YOLO OBB detection achieving 30Hz real-time inference through OpenVINO acceleration on Intel CPU/GPU/NPU hardware.
- Designed **automated synthetic dataset pipeline** using Gazebo Harmonic simulation with **TF2-based 3D→2D annotation projection**, eliminating manual labeling by leveraging ground-truth poses for oriented bounding box generation.
- Implemented **3D pose estimation** combining **pinhole camera model with ray-plane intersection** for world-frame position and orientation recovery, and executed full **ROS 2 Humble → Jazzy migration** including URDF, ros2_control, Gazebo Harmonic plugins, and MoveIt 2 configuration.

AI-Powered Rotary Water Meter Reading System, Technical Lead & Primary Author (64% of commits)

- Designed and implemented the core **multi-stage AI inference pipeline** chaining 5 distinct models: Geti SDK instance segmentation (meter validation), **YOLOv8 OBB detection** (meter ID + reading region localization), custom **OpenVINO OCR with CTC decoding**, and **RapidOCR fallback** with perspective warp rectification and **HSV-based rotation disambiguation**.
- Built smart post-processing layer with **fuzzy prefix repair**, context-aware OCR error correction (positional O↔0, 4↔A disambiguation), and **regex-based cleaning** that significantly improved field accuracy on real meter images.
- Delivered FastAPI inference server with async model lifecycle, PayloadCMS + PostgreSQL backend with **geolocation-tagged records**, and the complete training module (**YOLO OBB training, PaddleOCR dataset creation, PaddleOCR → OpenVINO model conversion**).

Visual-Textual Query-Driven Document Reasoning Engine, Sole Author

- Architected a **multi-modal agentic AI system** that enables natural-language querying of visually dense PDF documentation (schematics, diagrams, annotated photos) using ColQwen2.5 late-interaction retrieval and Qwen2.5-VL vision-language inference, eliminating reliance on OCR-based search.
- Implemented **OpenVINO-optimized VLM inference** (FP16/INT8/INT4 quantization) with automated model export, streaming response generation via **threaded TextIteratorStreamer**, and Intel XPU/GPU hardware acceleration verified on Arc A770 and Arrow Lake platforms.

- Built a **pluggable model registry** supporting 12+ model variants across 4 families, with domain-agnostic configuration enabling rapid adaptation to new use cases (robotic arm manuals, IPC-A-610 electronic standards) **without code changes**.
- Delivered production-ready Gradio web application with PDF upload, component image querying, retrieval gallery, persistent query history with **perceptual image deduplication**, and real-time token streaming for field technicians.

AI-Powered Robotic Arm: Vision-Language-Action (VLA) Framework, Primary Author

- Authored the complete **ROS 2 Humble driver** for the Toyo CHY2 industrial pneumatic gripper; service-based open/close control via TM5S SetIO, state machine with digital I/O pin feedback, joint state publishing, and auto-discovery of TM services, enabling end-effector control for all grasping tasks.
- Built a **real-time MyCobot-to-TM5S teleoperation bridge** for Learning from Demonstration, implementing both Python and C++ ROS 2 nodes with **configurable joint mapping, rate-limited command dispatch**, and gripper threshold control iterated from Python to C++ for **sub-millisecond processing latency**.
- Developed **TM Eye-in-Hand camera ROS 2 integration** streaming the robot's wrist-mounted camera via **gRPC** to standard ROS 2 Image topics, and integrated **Intel RealSense cameras** with **TF2 frame alignment** for multi-view policy training.

Real-Time Edge Speech Translation & Transcription, Sole Author

- Built **multi-threaded real-time speech translation system** running OpenAI Whisper (tiny-large-v3) optimized via **OpenVINO INT8/FP16 quantization** on Intel edge GPUs, supporting simultaneous translation to 3 languages (English, Mandarin, Japanese) with producer-consumer audio buffering for bounded latency.
- Designed two architectural variants, multi-pipeline (**3 parallel model instances**) for maximum throughput and single-pipeline (shared model with per-call generate_kwargs) for **~3x VRAM reduction** on constrained edge devices.

Agentic AI Manufacturing HMI; LLM-Driven PCB Defect Detection & Robotic Sorting,

Primary Author (79.5% of commits, sole author of all 16 core source files)

- Designed and built an end-to-end **agentic AI manufacturing system** integrating real-time computer vision, 6-DOF robotic automation, conveyor belt control, and an LLM-based natural-language interface for PCB defect detection and autonomous sorting on Intel edge hardware, demonstrated at **Embedded World 2025**.
- Implemented a **multi-frame defect detection pipeline** using MobileNetV2-ATSS (**OpenVINO FP16, 98.2% mAP**) with Intel Geti SDK, performing inference across 5 defect classes with **early-exit voting optimization** across up to 10 frames per PCB.
- Developed full-stack robotic automation, programmed **multi-waypoint pick-and-place** sequences for myCobot 280 with gripper actuation to sort PCBs into pass/fail bins, and engineered conveyor belt integration end-to-end from **Arduino** stepper motor firmware (**AccelStepper + serial command protocol**) to Python driver.
- Built **7 custom LLM tool-calling functions** for an OpenWebUI agent (Llama 3.2), enabling natural-language control of robotic inference, **real-time defect analytics** queries, and defect remediation knowledge retrieval via RAG, the differentiating feature bridging generative AI with physical automation.
- Architected the FastAPI backend serving **real-time dual MJPEG camera streams** (live + annotated), background-task inference orchestration, MongoDB-backed analytics engine with **time-windowed aggregation pipelines**, and automated chart generation (bar/line/pie); **8 REST endpoints** consumed by both the monitoring dashboard and the LLM agent.

OCS2 Optimal Control Framework (ROS 2 Migration)

- Ported 2 modules of a 138K-line C++ optimal control framework (OCS2, ETH Zurich) from ROS 1 to ROS 2 Humble, migrating build systems (catkin → ament_cmake), C++ APIs (roscpp → rclcpp), and resolving transitive dependency chains for Pinocchio/hpp-fcl/Eigen3.
- Migrated self-collision visualization system (81% commit ownership) rewriting MarkerArray-based RViz rendering for collision proximity debugging, used in MPC trajectory validation for manipulator and legged robot applications.