

KYAW SWAR HEIN

Data Scientist · Data Analyst · AI/ML Developer

heinm.dev@gmail.com | linkedin.com/in/kyawswarheinm | github.com/kyawswarheinm

Portfolio: kyawswarheinm.vercel.app | Pathum Thani, Thailand

PROFESSIONAL SUMMARY

Final-year ICT student and data science practitioner with two first-place competition awards and a published conference paper. Built end-to-end ML pipelines spanning automated data collection, NLP inference, computer vision model training, statistical analysis, and live dashboard deployment — covering the full workflow from raw data ingestion to actionable visualization. Seeking a Data Science internship where machine learning, NLP, and data analysis skills contribute to measurable business outcomes.

EDUCATION

B.Sc. Information and Communication Technology — Cumulative GPA: 3.96 / 4.00

Expected Graduation: Dec 2027

Rangsit University International College · Pathum Thani, Thailand

Relevant coursework: Big Data Analysis · Neural Networks & Deep Learning · Fundamentals of Machine Learning · Statistics · Database Management Systems · Object-Oriented Programming (Python, Java) · Internet of Things · Research Design · Project Management

TECHNICAL SKILLS

Languages — Python · SQL (PostgreSQL) · JavaScript · TypeScript · Java

ML & Data Science — Scikit-learn · TensorFlow · PyTorch · Pandas · NumPy · Feature Engineering · Model Evaluation · HuggingFace Transformers · NLP · Text Classification · Sentiment Analysis

Statistical Analysis — Linear Regression · Multiple Regression · Hypothesis Testing · Descriptive Statistics · Correlation Analysis · Time-Series Analysis

Computer Vision — Ultralytics YOLO · OpenCV · Object Detection · Video Stream Processing · Dataset Annotation

Frameworks & Tools — FastAPI · Flask · React · Node.js · Power BI · MongoDB · Git · Figma · Streamlit · Selenium · Plotly

Visualization — Chart.js · Matplotlib · Seaborn · Dashboard Design · KPI Reporting · Data Storytelling

PROJECTS

Customer Behavior Analysis with YOLOv26

Mar 2026

AI/ML · Computer Vision · Retail Analytics · Live demo: cba-prototype.vercel.app

- Trained a custom Ultralytics YOLO object-detection model on **12,578 annotated frames** from 11 real retail video clips across **4 interaction classes** (No Interaction, Touching Shelf, Holding Product, Item Removed), achieving **mAP@50 60.91%** · **Precision 72.69%** · **Recall 58.95%** — establishing a baseline for pre-purchase shelf behavior detection without a publicly available benchmark.
- Built a spatial tracking layer that mapped detected interactions to shelf zones and product locations, generating product-level engagement scores and uncovering high-attention, low-conversion areas, including frequently handled, returned, and abandoned products.
- Deployed a React + FastAPI analytics dashboard featuring **8 operational KPIs**, a dot-matrix dwell map, spatial interaction heatmaps, purchase-conversion funnels, and a **planogram-based interaction overlay**, with product-level engagement analysis and converting raw inference outputs into decision-ready retail intelligence.

Pantip SET Sentiment Monitor

May-Jul 2026

NLP · Automation · Sentiment Analysis · XLM-RoBERTa · Transformer Models · Live Demo: pantip-sentiment.streamlit.app

- Automated the **full data collection to visualization pipeline** — Selenium scraper (circuit breaker, 30s timeout) extracting Pantip.com investment threads → NULL timestamp recovery, JS-rendered reply count handling, and multi-ticker deduplication → 4-pass entity linking and XLM-RoBERTa batch inference → Turso cloud database → live Streamlit dashboard with sentiment rankings, trend charts, and yfinance price overlay — running end-to-end every 3 hours via GitHub Actions.
- Applied **cardiffnlp/twitter-xlm-roberta-base-sentiment** (multilingual XLM-RoBERTa) to **2,500+ Thai-language** Pantip.com investment threads linked to **50 SET-listed tickers**, producing per-post sentiment scores (−1 to +1) with a **0.65 confidence cutoff** (~2× random-chance baseline for a 3-class model), storing label, sentiment, and confidence per prediction.
- Ran **Pearson/Spearman lag-correlation backtest** across lags 0–5 days between daily sentiment aggregates and yfinance closing-price returns to test whether Thai retail forum sentiment predicts SET stock movement; implemented z-score ($\geq 2.5\sigma$) and volume-surge (3× 7-day avg) anomaly detection to flag statistically significant sentiment spikes.
- Built a **4-pass entity matcher** (exact symbol, company name, alias dictionary, RapidFuzz fuzzy at threshold 92) and **automated the full pipeline** every 3 hours via GitHub Actions; deployed a live Streamlit dashboard and published the scored dataset on Kaggle, achieving a **10.00/10.00 usability score**.

CiMSO Hospitality ERP Analytics Dashboard — 1st Place, Hackathon

Mar 2025

Data Visualization · Dashboard Dev · AI Integration · Live demo: hotel-eternora-dashboard.vercel.app

- Won 1st Place **out of 11 finalist teams** at the CiMSO Hospitality ERP Hackathon 2025, judged on analytical depth, design clarity, and presentation quality — competing against teams from across the university.
- Architected a hospitality analytics platform with **25 Chart.js visualizations and 45 KPIs** across **9 operational categories** (booking arrivals, members and guests, arrivals and departures, occupancy and ADR, guest birthdays, age groups, booking statuses, unit information, and income) with time-period and category-based filtering, color-blind-friendly palettes, and raw-data tables view for record-level audit.
- Integrated an LLM-powered insight assistant supporting **text-to-speech and speech-to-text input across 5 languages** (English, Thai, Malay, Traditional Chinese, Burmese), broadening accessibility for multinational hotel management teams and achieving full translation coverage of both static UI and AI-generated analytical content.

Conference Research Paper — Published at RSUCON 2026

May 2026

Statistical Analysis · Python · Linear Regression · ISBN (e-book) 978-616-421-246-6, pp. 387–397

- Co-authored and published a peer-reviewed quantitative study in the **RSUCON 2026 Proceedings** (international research conference), establishing a citable academic publication record.
- Executed Python-based multiple linear regression on n=196 student responses across 3 independent variables (Digital Multitasking, Cognitive Load, Technology Use); isolated Technology Use as the sole significant GPA predictor ($\beta=0.137$, $p=0.004$, Adjusted $R^2=0.031$), disconfirming the multitasking-performance hypothesis at $\alpha=0.05$.
- Presented research findings in an in-person academic Q&A session, defended statistical methodology against faculty reviewers, and delivered structured rebuttal to cross-discipline questions — strengthening skills in evidence-based communication.

System Design: Modern Library Management System

Apr 2025

System Analysis & Design · Context Diagram · DFD · ERD · Use Case

- Produced a complete software architecture case study — **context diagram, two-level DFDs, normalized ERD, and use-case diagrams** — modeling borrowing, returns, overdue handling, and inventory workflows across multiple actor roles (student, librarian, university registrar).
- Applied structured systems-analysis methodology to decompose domain requirements into process and data models — producing artifacts directly transferable to ML pipeline documentation, data schema design, and system architecture planning.

EXPERIENCE & PRESENTATIONS

IoT Workshop Presenter & Demonstrator — Rangsit University

Apr 2026

Python · NodeMCU · HC-SR04 Ultrasonic · Servo Control · Real-time Visualization

- Built and demonstrated a real-time 180° ultrasonic sonar scanner (NodeMCU + HC-SR04 + SG90 servo) streaming distance-angle data to a **live Python sonar visualization**, showcasing an end-to-end IoT sensor-to-display pipeline to workshop participants.
- Delivered technical instruction on PWM servo control, ultrasonic ranging physics, and serial data streaming to an audience with mixed hardware backgrounds, achieving **full session completion** across all hands-on participants.

Project Showcase Presenter — RSU Job Fair 2026

Feb 2026

Public demo to recruiters, students, and university leadership including RSU President Dr. Attawit Ourairat

- Represented the award-winning CiMSO ERP analytics platform across a **12-hour public exhibition**, delivering tailored technical demonstrations to recruiters, students, faculty, and university leadership — adapting depth and framing to each audience type in real time.
- Fielded unsolicited technical and business questions from industry recruiters regarding **AI integration, dashboard architecture, and multilingual accessibility** — demonstrating the ability to articulate data engineering decisions in non-technical terms under professional scrutiny.

AWARDS

🏆 1st Place — CiMSO Hospitality ERP Hackathon 2025

Mar 2025

- Outperformed 11 competing finalist teams; judged on analytical depth, design quality, and presentation.

🏆 1st Place — RIC Startup & Entrepreneurship Pitching Sessions

Jul 2025

- *LÉOR* wearable safety startup; led narrative design, Figma prototype, and pitch delivery

CERTIFICATIONS

Statistics for Data Science with Python · IBM / Coursera

July 2026

IBM Data Science · IBM / Coursera

Jun 2026

Google AI Essentials · Google / Coursera

Jun 2026

Odoo Education Workshop (APAC) · Odoo HK Limited

Feb 2026