

# YOHANES ANDRE SETIAWAN

[mail@andre44s.dev](mailto:mail@andre44s.dev) | [linkedin.com/in/andre44s](https://linkedin.com/in/andre44s) | [Bali, Indonesia](#)

AI/ML engineer with four years of experience across education, e-commerce, and mining. I build production systems for LLMs and agentic AI, RAG/retrieval, recommendation, NLP, and computer vision, and ship end-to-end ML pipelines (training, evaluation, deployment/inference). Proficient with Python, PyTorch/TensorFlow, SQL, MLOps, and cloud (AWS, Azure, GCP), including Docker, vector search/embeddings, and CI/CD. M.S. in Computer Science (AI) from Georgia Institute of Technology.

## EXPERIENCE

---

### **Datamine — Senior AI Engineer**

January 2025 – Present

- Engineer an LLM-powered RAG system (vector search + retrieval) for semantic search over product docs to replace keyword search.
- Develop an MCP-based agent to automate multi-step workflows and reduce manual effort.
- Design a multi-agent system that generates proprietary code to accelerate the SDLC.
- Architect Azure CI/CD and MLOps for deployment/inference of AI services in production.
- Integrate AI services into C#/.NET and REST APIs for production use.

### **Tokopedia — Data Scientist**

October 2023 – December 2024

- Implemented a two-tower retrieval/ranking model to improve recommendation.
- Engineered product deduplication and diversification using clustering algorithms to diversify feeds and reduce duplication.
- Integrated user negative-feedback into the recommender to reduce irrelevant results.
- Improved age prediction accuracy by 5% using ensemble techniques, with an 80% reduction in feature reliance.
- Prepared training datasets in GCP BigQuery with SQL to support model development.
- Maintained the complementary recommendations to support cross-sell use cases.
- Built Airflow DAGs orchestrating C++ components for model training, evaluation, and deployment workflows.

### **CoLearn — Data Scientist**

September 2021 – October 2023

- Designed a recommender system employing PyTorch transformers and FastAPI for personalized video delivery.
- Built PoCs with OCR, NLP, and LLM prompting for a student teaching assistant.
- Performed data exploration and cleaning on user behavior with AWS Athena and Python.

- Created GenAI/LLM solutions (RAG/retrieval, evaluation) using LangChain and Hugging Face.

## EDUCATION

---

### Georgia Institute of Technology

January 2024 – May 2026

#### *Master's in Computer Science (AI)*

- Specialization in AI with applications in variety of fields, including robotics, finance, gaming, and computer architecture. Current GPA: **3.87/4.00**.
- Coursework: Machine Learning for Trading, AI for Robotics, High Performance Computer Architecture, Game AI, Machine Learning, Deep Learning, Knowledge-Based AI, Network Science, Software Development Process, GPU Hardware and Software.

### Udayana University

June 2018 – October 2022

#### *Bachelor's in Electrical Engineering (Electronic, Robotic, and Control System)*

- Head of the Robotics Club, managing 100+ members for the Indonesia Robot Contest.
- GPA: **3.88/4.00**, culminating in a thesis on Computer Vision and Neural Networks in Robotics.

## SKILLS

---

### ML/AI

- Recommender systems (two-tower retrieval/ranking), Transformers, Time series.
- RAG (semantic search + retrieval), LLMs and AI agents (MCP).
- OCR, text classification and extraction, computer vision.
- LLMs and GenAI (prompting, evaluation, deployment/inference), vector search/embeddings.

### Frameworks and Libraries

- Python, SQL, C/C++.
- PyTorch, TensorFlow, scikit-learn, LangChain, Hugging Face, Pandas.
- FastAPI, REST design, integration with C#/NET production systems.

### MLOps and DevOps

- CI/CD on Azure (pipelines, release automation), MLOps, Airflow, testing and monitoring basics.
- Containerization and orchestration: Docker, Kubernetes; model deployment/inference.
- Microsoft Azure, Amazon Web Services (AWS), and Google Cloud (GCP).

## OTHER

---

- **Language:** Indonesian (Native), English (Professional – C1)
- **Certification:** Certified TensorFlow Developer (2021), AWS Certified Solutions Architect – Associate (2023)