

# Dzaky Zakiyal Fawwaz, Ph.D.

AI/ML Engineer | AI Agents, LLMs & RAG | Deep Learning & Computer Vision | Team Lead

Jakarta, Indonesia (UTC+7), Open to Remote | +62 851 2110 8575 | [dzakybd@gmail.com](mailto:dzakybd@gmail.com)

LinkedIn: [linkedin.com/in/dzakyzf](https://www.linkedin.com/in/dzakyzf) | GitHub: [github.com/dzakybd](https://github.com/dzakybd) | Portfolio: [dzaky.ai](https://dzaky.ai)

## PROFESSIONAL SUMMARY

AI/ML Engineer with a Ph.D. in Computer Engineering and 7+ years of experience across both Generative AI and traditional machine learning. Currently ships AI agents to a production GenAI platform used by global brands, builds self-hosted voice AI (STT, LLM, TTS) on AWS GPUs, and leads a Data & AI stream delivering deep learning models and an agentic workflow for remote sensing applications. Equally hands-on with LLM systems (Pydantic AI, LangGraph, Deep Agents, RAG, MCP, LLM evaluation) and classical ML/DL (PyTorch computer vision, XGBoost, clustering, optimization, reinforcement learning), from data pipelines on Databricks and Spark Hadoop to deployment on AWS. Brings 3+ years of remote work with distributed international teams, plus 5+ years of AI research in South Korea.

## SKILLS

**Generative AI & Agents:** Generative AI, AI Agents, Multi-Agent Systems, Deep Agents, Pydantic AI, LangGraph, LangChain, RAG, MCP, Prompt Engineering, LLM Evaluation (pydantic-evals, LLM-as-judge), Langfuse, pgvector, BM25 hybrid search

**LLM & Voice AI:** vLLM, TensorRT-LLM, Qwen, OpenAI / Azure OpenAI, Gemini / Vertex AI, OpenRouter, Pipecat, Whisper (STT), GPT-SoVITS (TTS fine-tuning), Silero VAD, WebRTC

**Machine Learning & Deep Learning:** PyTorch, Computer Vision, Semantic Segmentation, Self-Supervised Learning, XGBoost, Scikit-learn, Reinforcement Learning, Integer Programming, Genetic Algorithms, NLP, OCR (PaddleOCR, Docling), Hugging Face.

**Data & MLOps:** Python, SQL, Databricks, MLflow, PySpark, Apache Sedona, Airflow, Kubeflow, Remote Sensing, GeoPandas, rasterio, H3.

**Cloud & Engineering:** AWS (ECS, EKS, S3, ECR), GCP (Cloud Run, Compute Engine), Azure (App Service, DevOps), Terraform, Docker, GitHub Actions, FastAPI, REST, GraphQL, OAuth 2.0, PostgreSQL, Prometheus.

**Languages:** English (Professional, C1), Indonesian (Native)

## PROFESSIONAL EXPERIENCE

**Accenture**, Jakarta, Indonesia

06/2024 – Present

*AI Decision Science Consultant (Industry & Functional AI) | Full-time*

**Data & AI Stream Lead** | Satellite Methane Detection Platform, Oil & Gas Client

07/2026 – Present

- Lead the Data & AI stream for a methane detection platform going to production in October 2026, fusing Sentinel-2, Landsat 8/9, and EMIT satellite imagery with ground gas-sensor (FGDS) readings.
- Designed the hybrid on-premise/AWS architecture: Airflow ingestion, PySpark/Apache Sedona medallion pipelines, Dataiku GPU training on AWS EKS, and Databricks agents with Lakeflow Jobs publishing to ArcGIS.
- Trained PyTorch semantic segmentation models (U-Net, DeepLabV3+ with ConvNeXt, SimCLR self-supervised pretraining on 57K tiles), reaching an F1 above 0.7 on real plumes and estimating emission flux with 5% mean error after calibration.
- Built a Deep Agent (LangGraph) that correlates detections with sensor, asset, work-order, P&ID, and HSE data to find root cause, recommend next-best actions, and write cited reports, evaluated via LLM-as-judge and traced in Langfuse.

**AI Engineer** | Conversational Voice AI Platform, Telco Client

2026 – Present

- Building a self-hosted, real-time voice agent (STT → LLM → TTS) on Pipecat with WebRTC, Silero VAD, a pgvector RAG knowledge base, and 6-language switching (Indonesian, English, Japanese, Korean, Chinese, Cantonese); telephony PoCs on Asterisk SIP and WhatsApp calling.
- Fine-tuned GPT-SoVITS TTS for Indonesian with a custom phoneme (G2P) frontend; served Whisper on TensorRT-LLM and Qwen3-32B (AWQ) on vLLM.
- Deployed GPU services on AWS ECS with Terraform, GPU-metric autoscaling, and CloudWatch; load-tested STT, LLM, and TTS latency and throughput with Locust.
- Evaluate and integrate emerging voice, multimodal, and AI avatar capabilities (voice cloning, lip-synced avatars) into reusable AI assets.

**AI Engineer** | GenAI Ad Creation Platform, production SaaS (Germany-based team, remote)

09/2025 – Present

- Shipped AI agents (Campaign Brief, Template, and Creative Agents) built on Pydantic AI Deep Agent (Agent Harness) to the Python/FastAPI AI orchestration service of a multi-tenant production platform used by global consumer brands, with human confirmation before sensitive actions.

- Built a RAG pipeline (document processing & retrieval) for the agents, applying chunking, embedding, and pre- and post-retrieval techniques to understand brand guideline context from various files using Docling and pgvector.
- Created the LLM evaluation suite (pydantic-evals, LLM-as-judge) and a GitHub Actions gate that blocks PRs on quality regressions; added reasoning-token streaming (SSE), run cancellation, and multi-question tools.
- Integrated computer vision models (LayerD, LaMa inpainting, OCR) to split ad images into editable layers.

**AI Governance Consultant** | Responsible AI, Oil & Gas Client 03/2025 – 06/2025

- Worked with client stakeholders and teams across countries to deliver an AI technology landscape assessment, an AI technology selection guide (LLM benchmarking, RAG vs fine-tuning), and AI governance risk frameworks.
- Built a Responsible AI testing toolkit (Streamlit, Ragas, Presidio, Detoxify, LLM Guard) and assessed production ML and RAG use cases for toxicity, PII, bias, explainability (SHAP), and answer correctness.

**Data Scientist** | Retail Store Network Optimization, Telco Client 10/2024 – 01/2025

- Contributed to an XGBoost visit-demand model ( $R^2$  0.845 on 547K H3 grid cells) and network optimization for 478 stores using integer linear programming (PuLP) and a multi-objective genetic algorithm.
- Built K-means visitor segmentation on PySpark and LLM-generated per-store justifications on Kubeflow and MLflow; recommended scenarios identified IDR 17.6–48.5B (about US\$1.1–3.0M) in annual cost savings.

**AI Engineer** | Customer Segmentation & GenAI Video Analysis, FMCG Client 08/2024 – 10/2024

- Built containerized RFM segmentation (77 K-means models across brands and marketplaces) and data-cleansing pipelines covering 5.8M customers.
- Contributed to a GenAI video-understanding solution that analyzes affiliate videos in minutes, not hours of manual review.

**Delman Data Lab**, Indonesia (remote) 03/2022 – 04/2024

*Data Scientist | Full-time*

- Led FMCG reporting automation, cutting operational costs by 50% and processing time from hours to minutes.
- Led procurement data standardization for a manufacturer, identifying 50% potential cost savings.

**Daewoong Pharmaceutical**, South Korea (remote) 07/2021 – 09/2022

*AI Engineer | Part-time*

- Built a real-time computer vision model for pharmaceutical droplet encapsulation inspection (94% accuracy).
- Built an AI health assistant chatbot with medication reminders, drug identification, and recommendations.

**INSLAB, Pusan National University**, South Korea 09/2018 – 02/2024

*AI-IoT Researcher (Graduate Research) | Full-time*

- Improved the 6TiSCH IoT protocol with multi-agent Q-learning: +12% packet delivery, -23% energy, -9% latency.
- Built hydraulic system fault detection with an evolutionary algorithm and autoencoder (~98% accuracy).

## Earlier Experience

- **Software Developer, MUFKO, South Korea (01/2019 – 12/2019)**: GraphQL backend with Google OAuth login for a halal app (100K+ downloads); also contributed to a Laravel internal tool and the Flutter frontend.
- **AI Research Intern, CITI – NTUST, Taiwan (07/2018)**: OpenPose + LSTM activity recognition (98%) and OCR.
- **Software Developer Intern, Chevron, Indonesia (07/2016 – 01/2017)**: drilling data reporting automation.

## EDUCATION

**Ph.D. (Integrated M.S./Ph.D.) in Computer Engineering**, Pusan National University, South Korea 2018 – 2024

GPA 4.28/4.50 | Thesis: Improving 6TiSCH Network Formation and Transmission using Q-Learning

**B.S. in Computer Science**, Institut Teknologi Sepuluh Nopember (ITS), Indonesia 2014 – 2018

GPA 3.72/4.00 | Thesis: Human Activity Recognition on Video Data using 3D-CNN

## SELECTED PUBLICATIONS

8 peer-reviewed papers (3x IEEE Access, Sensors, Applied Sciences, IEEE ICOIN) on reinforcement learning and edge AI for IoT. Google Scholar: [scholar.google.com/citations?user=AhqpfOQAAAAJ](https://scholar.google.com/citations?user=AhqpfOQAAAAJ)