BAYU DIAN NUGROHO

Senior Software Engineer | AI Solutions Architect

Indonesia • +62-822-2822-0181 • kleirose.bayu@gmail.com • linkedin.com/in/bayu-dian-nugroho • github.com/rhythmwave

SUMMARY

Senior Software Engineer with 10+ years of experience designing cost-aware, production-grade AI systems. Specialized in building multi-tenant platforms that scale to zero when idle, minimizing infrastructure costs while maintaining enterprise-grade reliability. Led architecture and implementation of AI platforms serving enterprise clients with strict security and compliance requirements. Strong focus on Kubernetes-native architectures, event-driven systems, and observability-driven operations.

PROFESSIONAL EXPERIENCE

Senior Software Engineer / AI Engineer | *PT Berlian Sistem Informasi* **May 2024 – Present**

Led the architecture, backend engineering, and cost-optimization strategies for enterprise-grade distributed services and multi-tenant AI platforms. Spearheaded multiple production-ready AI products from infrastructure design to feature implementation:

- **Enterprise AI Platform (EIRA):** Architected a modular microservices platform on Azure Kubernetes Service (AKS) featuring Microsoft Entra multi-tenant authentication, delivering 6 production-ready features (Document QnA, Translation, MoM Extraction) while reducing manual meeting documentation by 70% and maintaining 100% data privacy compliance.
- **WhatsApp AI Assistant Platform (DESTA):** Designed a single-webhook NLP routing and RAG-powered response system integrated with Twilio, implementing defense-in-depth security (signature verification, request replay prevention) to reduce dealership support response times by 60% with >95% intent classification accuracy.
- **Immersive Role-Play Platform (AI Trainer):** Launched a real-time sales training platform utilizing WebSockets, Rive animations, and ElevenLabs voice synthesis, incorporating a secure, backend-provisioned token system for cost control that effectively minimized trainer dependency for over 100+ active salesmen weekly.
- **Customer Intelligence Platform (PTC):** Engineered an event-driven async workflow utilizing Azure OpenAI Transcribe and Diarization to process service entry conversations, delivering pre-service customer insights via Flutter mobile apps and reducing Service Advisor preparation time by 50%.
- **Core Infrastructure & Cost-Optimization:** Implemented scale-to-zero automation via KEDA and centralized shared LLM gateways (LiteLLM), establishing shared Azure OpenAI rate limiting to drastically lower per-tenant compute and operational overhead.
- **High-Performance Services:** Built a distributed chunked upload service in Go backed by Redis coordination, and developed a Python-based document extraction engine that accelerated processing speeds by 85% (from ~3 minutes down to 10–30 seconds).
- **Search Optimization:** Developed proprietary TF/IDF and fuzzy-search pipelines that achieved near-embedding retrieval quality compared to commercial vector databases.
- **Observability Engineering:** Built a high-throughput Go-based observability gateway using NSQ and Fasthttp, successfully streaming logs, metrics, and traces into Loki, Tempo, and Mimir for comprehensive AI pipeline health monitoring.
- **Drone AI Inspection Platform:** Architected an end-to-end AI-powered drone inspection system for large-scale corporate vehicle facilities (10K+ vehicles), enabling automated detection of bird droppings and surface debris. Designed the complete AI processing pipeline covering drone data ingestion, image processing, distributed inference, and real-time field operations.
- Built computer vision workflows using YOLO-based object detection, SAHI optimization, and transformer-based vision models, trained on 5,000+ labeled images using Azure GPU infrastructure to improve detection accuracy on large-area inspection scenarios
    - Key Achievements: Led architecture and delivery of multiple production AI platforms across enterprise workflows, solving challenges in document intelligence, conversational AI, voice-based training, customer analytics, and computer vision inspection. Improved operational efficiency through AI automation, reduced processing latency by up to 85%, and optimized cloud infrastructure cost through Kubernetes-based scale-to-zero architecture.

Software Engineer | *SCCIC ITB* Sep 2023 – Apr 2024

- Integrated YOLO-based face and object detection into backend services for production video analytics workloads.
- Deployed services across cloud and on-prem environments using Docker, Airflow, and Nginx with automated workflows.
- Built backend services in Go and contributed across Node.js, Python, and PHP stacks in small engineering teams.

- Key Achievements: Delivered stable, multi-environment service rollouts and production deployment of ML-powered video analytics.

DevOps Engineer | *PT Neuronworks Indonesia* Jan 2023 – Sep 2023

- Implemented Kubernetes-based orchestration to reduce infrastructure costs and improve deployment reliability.
- Containerized legacy and new applications with Docker and set up automated GitLab CI/CD pipelines.
- Configured enterprise monitoring and alerting infrastructure with Prometheus, Grafana, and related observability tools.
  - Key Achievements: Successfully reduced overall infrastructure operational costs through containerization and automated orchestration.

Junior System Analyst | *PT Neuronworks Indonesia* Jan 2020 – Jan 2023

- Analyzed, modeled, and thoroughly documented business and technical system requirements across multiple enterprise projects.
- Led the technical migration of monolithic applications into microservices architectures using Docker, Kubernetes, and MinIO-backed storage.
- Implemented geolocation-based fraud prevention and security verification features for financial applications.
  - Key Achievements: Ensured successful migration from monolith to microservices, vastly improving system modularity and deployment flexibility.

Software Developer | *PT Neuronworks Indonesia* Mar 2013 – Dec 2019

- Developed robust web applications using PHP and Zend Framework tailored for enterprise clients.
- Implemented ETL-based data analytics pipelines to generate business intelligence and reporting dashboards.
- Contributed heavily to technical documentation and internal code standardization efforts, establishing the team's core coding guidelines.

E D U C A T I O N

ST INTEN | *Bachelor Degree, Information Systems* 2011 – 2015

- Developed strong foundations in information systems, relational databases, and core software engineering principles.
- Built early web applications and internal utilities as part of academic coursework and side projects.

T E C H N I C A L S K I L L S

Backend Engineering: Go (Advanced), Python (Advanced), PHP (Advanced), Node.js (Intermediate), C#/.NET (Intermediate), C/C++ (Intermediate), REST APIs, WebSocket, gRPC, Microservices Architecture

Cloud & DevOps: AWS (Intermediate), Azure (Intermediate), GCP (Intermediate), Tencent Cloud (Intermediate), Kubernetes (Intermediate), Docker (Advanced), Podman (Intermediate), KEDA, ArgoCD, Jenkins, GitHub/GitLab CI, Terraform, Linux (Advanced)

Data & Messaging: PostgreSQL (Advanced), MySQL (Advanced), CockroachDB, MongoDB (Intermediate), OracleDB (Intermediate), Vector Database (Intermediate), MinIO, Redis, Kafka (Intermediate), NSQ, RabbitMQ (Intermediate)

AI / ML Systems: AI Agents, LLM Integration (OpenAI, Azure, Google, Anthropic), Prompt Engineering, RAG Systems, LangChain, Azure AI Search, LiteLLM Gateway, YOLOv8, OpenCV, Document Intelligence (PDF/Office), TF/IDF, Fuzzy Search, ElevenLabs Voice Synthesis

Real-Time & Integration: WebSocket, Rive Animation, Twilio/WhatsApp API, Moodle LMS Plugin Development, Microsoft Entra (Multi-tenant SSO), Azure Transcribe + Diarization, Flutter (Intermediate)

Observability & Tools: Grafana (Intermediate), Prometheus/Mimir (Intermediate), Loki (Intermediate), Tempo (Intermediate), OpenTelemetry (Intermediate), Git (Advanced)

System Design Patterns: Distributed Systems, Cost-Aware Architecture, Scale-to-Zero Architectures, Event-Driven Architecture, Multi-tenant Identity Management, Plugin-based System Integration, Microservices Design

Frontend: JavaScript (Advanced), React (Intermediate), Vue (Intermediate), HTML5 (Advanced), CSS3 (Advanced), Tailwind CSS (Advanced), Bootstrap (Advanced)

L A N G U A G E S

- Indonesian (Native)
- English (Professional Working Proficiency)