

# RIZELLE B. BAUTISTA

rizellebautista123@gmail.com | (+63) 916-406-6931 | Cainta, Philippines | [LinkedIn](#) | [Portfolio](#) | [GitHub](#)  
[AWS Developer Associate](#) • [AWS AI Practitioner](#) • [AWS Cloud Practitioner](#) • [CompTIA ITF+](#)

---

## PROFESSIONAL SUMMARY

AI Systems Engineer with proven production experience designing, building, and shipping agent-based and conversational AI systems end-to-end. Built a production GenAI platform at WebPuppies (Singapore) using Amazon Bedrock, LangChain, and agentic workflows — live in production today. Architected Lily, a fully autonomous conversational AI assistant integrating five simultaneous AI pipelines — RAG, LLMs, emotion detection, speech-to-text, and text-to-speech — with human-oversight guardrails and automatic cloud-to-local failover on edge hardware. Experienced in designing human-in-the-loop systems, building evaluation and feedback frameworks, and implementing hallucination guardrails for AI outputs in high-trust environments. Strong Python engineer comfortable with LangGraph agentic state management, persistent memory, tool-use patterns, and production AI service deployment. Mission-driven — specifically motivated by applying AI to improve real human outcomes at scale.

---

## PROFESSIONAL EXPERIENCE

**Software Engineer — AI Systems** | *WebPuppies, Singapore (Remote)* Dec 2025 – Present

**Pioneer Lead — Agentic AI Platform, LLM Workflows & Human-in-the-Loop Systems**

- Designed and shipped a production agentic AI platform from scratch — translating stakeholder business requirements into a scalable system using Amazon Bedrock (Claude), AWS Lambda, Step Functions, and API Gateway — now live on the company website and central to their AWS APN partnership.
- Built human-in-the-loop safeguards across all AI outputs by implementing Amazon Bedrock Guardrails, iterative LLM refinement loops, and structured output validation — reducing hallucinations to near-zero on enterprise financial reports and ensuring every AI-generated output met human review standards before client delivery.
- Designed and built evaluation infrastructure for AI output quality — implementing iterative refinement loops where each output is validated against defined criteria, with automatic re-generation on failure — a production AI quality assurance framework directly applicable to high-trust care delivery environments.
- Engineered scalable agentic workflows on AWS Step Functions processing 120,000+ row datasets per run with real-time status updates, automated error handling, and retry logic — demonstrating production-grade agent reliability at enterprise scale.
- Collaborated directly with the Head of Technology throughout the product lifecycle — translating nuanced business requirements into AI system designs, presenting progress at each milestone, and securing architectural sign-off.

**President** | *AWS Learning Club — JRU Chapter* Mar 2025 – Present

- Led and mentored a 150+ member technical organization — conducting structured AI and cloud sessions, guiding 15 members to AWS certification, and translating complex AI concepts clearly for audiences ranging from complete beginners to certification candidates.
- Co-organized AWS Student Community Day Mega Manila 2025 drawing 400+ participants, managing cross-team Agile delivery via Jira.

**Web Developer & Data Analyst Intern** | *Analytics & AI Association of the Philippines* Oct 2024 – Present

- Built and shipped two production full-stack websites (aap.ph, aicon.ph) with React/Next.js, third-party API integration, and full operational ownership post-launch including data analysis and stakeholder-facing strategic reports.
- 

## KEY AI SYSTEMS PROJECTS

**Lily — Conversational AI Assistant with Emotion Detection & Human Oversight** | *Python, FastAPI, LangChain, Faster-Whisper, ElevenLabs, CNN, FAISS, pgvector, AWS, Raspberry Pi 5* 2024–2025

**Undergraduate Thesis | Project Manager, AI Systems Engineer & Full-Stack Developer**

- Architected and built a fully autonomous conversational AI system integrating five simultaneous AI pipelines on edge hardware (Raspberry Pi 5): Faster-Whisper STT with DSP noise reduction for multilingual speech recognition, CNN-based emotion detection (85% accuracy) to dynamically adapt responses to user emotional

state, Hybrid Adaptive RAG for grounded knowledge retrieval, dual-mode LLM (Groq Llama 3.1-8B cloud / Ollama Llama 3.2:1b local), and dual TTS (ElevenLabs / pyttsx3) — directly analogous to building a conversational agent for real-world care delivery.

- Designed and implemented a human-in-the-loop content validation layer — using Groq as an AI Content Validator to verify every RAG response against source documents before delivery, ensuring factual accuracy and reducing hallucinations in student-facing interactions — the same trust and safety pattern required in mental health AI systems.
- Built a Hybrid RAG architecture with seamless online-to-offline failover — when internet connectivity dropped, the system transparently switched to a fully local LLM and TTS stack with zero service interruption — demonstrating resilient AI service design for environments where reliability is non-negotiable.
- Integrated emotion detection into the conversational pipeline to modulate agent responses based on detected user emotional state — a direct analogue to affective AI design for mental health support contexts.
- Maintained 2–4 second end-to-end response times under continuous production load for 1,000+ projected users, while orchestrating five simultaneous AI services on constrained edge hardware — demonstrating production AI performance engineering.
- Reduced solution cost 80% versus industry-standard campus robots by leveraging hybrid cloud-edge architecture — enabling enterprise-grade AI delivery at accessible cost, relevant to mission-driven healthcare startups operating under resource constraints.

### **LangChain Conversational Chatbot with Persistent Memory** | *Python, LangChain, LangGraph, React* — *LangChain Academy Capstone* 2024

- Built a production multi-turn conversational AI agent using LangChain orchestration, LangGraph for agentic state management, and persistent cross-session memory — demonstrating deep mastery of the exact frameworks used in agent-based conversational systems.
- Implemented prompt templating, output parsers, memory buffers, and tool-use patterns using LangChain Expression Language (LCEL) — directly transferable to building agent-based care support workflows with context retention across sessions.

### **MedRoute AI — Emergency Healthcare Routing with Voice I/O** | *Python, ElevenLabs TTS/STT, Geospatial AI, Databricks* — *HackNation AI 2026* 2025

- Built an AI-powered emergency routing engine reducing patient-to-facility matching from 20+ minutes to under 3 seconds using natural language voice intake — designing AI for a high-stakes, human-safety-critical domain where errors have direct consequences for real people.
- Conducted geospatial analysis identifying 12 medical deserts and 34 critical equipment gaps — translating AI-generated insights into actionable findings for non-technical healthcare stakeholders, demonstrating the clinical-to-technical translation skill this role requires.

### **THRIVE — Inclusive Job Platform with AI Matching & Accessibility** | *React Native, Supabase, PostgreSQL, Python, STT/TTS* — *PLDT Season 4 Finalist* 2024

- Team Leader and Full-Stack Developer for an AI-powered inclusive job platform serving people with disabilities — integrating NLP-based job role matching, STT/TTS accessibility features validated by PDAO Cainta, and role-based dashboards; conducted iterative user testing with PWD communities across multiple regions.

### **AWS Bedrock AI Cost Consultant — Evaluation & Refinement Framework** | *Amazon Bedrock, Claude, Python, Node.js* 2024

- Designed a multi-step AI evaluation and refinement pipeline within Amazon Bedrock — each output is validated against business rules, and failures trigger re-generation with the failure reason injected as context — a production AI quality measurement framework directly applicable to building AI evaluation infrastructure.

---

## **TECHNICAL SKILLS**

**Agentic & Conversational AI:** LangChain (Academy certified), LangGraph, Bedrock Agents, multi-agent orchestration, agentic state management, tool-use patterns, persistent memory, cross-session context retention

**LLMs & AI APIs:** Claude (Anthropic), Amazon Bedrock, Groq Llama 3.1/3.2, Ollama (local), OpenAI, Gemini — prompt engineering, iterative refinement, hallucination reduction, structured output

**RAG & Knowledge Systems:** Hybrid Adaptive RAG (FAISS vector search + pgvector), document chunking and embedding, similarity search, AI content validation layer, knowledge base management

**Human-in-the-Loop & Safety:** Amazon Bedrock Guardrails, AI content validators, output evaluation frameworks, iterative refinement loops, responsible AI practices in high-trust environments

**Voice & Speech (ASR/TTS):** Faster-Whisper STT (multilingual: English + Tagalog) with DSP noise reduction, ElevenLabs TTS, pyttsx3, Azure Cognitive Services STT/TTS — production voice pipeline design

**Evaluation & Experimentation:** AI output quality frameworks, iterative refinement loops with failure injection, benchmarking pipelines, CloudWatch-based AI service monitoring, production observability

**Backend & APIs:** Python (Expert), FastAPI, Node.js, RESTful APIs, Microservices, async processing — production AI service design and deployment

**AI Infrastructure:** AWS Lambda, Step Functions, API Gateway, S3, DynamoDB, CloudWatch, IAM, Secrets Manager, Bedrock, SAM CLI, CodeBuild, CodePipeline — production AI platform operations

**Data & Databases:** PostgreSQL, MySQL, DynamoDB, Supabase, Databricks (Delta Lake, Unity Catalog), FAISS vector store, pgvector, SQL — strong SQL fundamentals

**AI Tooling:** Cursor, Claude Code, GitHub Copilot — daily AI-assisted development workflow

**Frontend:** React, Next.js, React Native, Streamlit — production UIs integrated with AI backend services

**Project & Collaboration:** Jira (Agile/Scrum), SDLC, technical documentation, cross-functional and cross-border team collaboration

---

## CERTIFICATIONS

- AWS Certified Developer – Associate
  - AWS Certified AI Practitioner
  - AWS Certified Cloud Practitioner
  - LangChain Academy — LangChain & LangGraph Curriculum
  - CompTIA ITF+
- 

## AWARDS & RECOGNITION

- Hack-Nation AI Global Hackathon 2026 — Selected alongside participants from MIT, Harvard, and Stanford University
  - 2nd Place — Globe Telecom Code-To-Career Hackathon 2025 | Python data engineering for real-world solution development
  - Finalist — PLDT & Smart Innovation Generation Season 4 | Led full-stack development of THRIVE
  - Super AI 2025 Singapore Delegate — One of 100 students worldwide selected to attend the largest international AI summit
- 

## EDUCATION

**Bachelor of Science in Computer Engineering** | *Jose Rizal University, Philippines*      2022 – 2026

*Dean's Lister • DOST SEI Scholar*