

Gedy Palomino

Senior Cloud & AI Engineer

email: gedy.palomino@gmail.com

phone: +51 902729770

Summary

Senior software engineer with 6+ years designing, building, and operating end-to-end systems. Strong background in backend architecture, infrastructure, and event-driven distributed systems. Hands-on experience with managed services and Linux system administration. Currently focused on AI-powered applications, automation, and agent-based development workflows. Solid working knowledge of Linux, networking, SQL/NoSQL data systems, authentication, cryptography, and security engineering.

Technical Skills

AI and developer tooling: Tool-using LLM agents, RAG workflows, MCP integrations, Structured outputs and schemas, Prompt and context engineering, LLM inference and benchmarking, AI automation pipelines, AI security and risk mitigation, AI observability

Backend and cloud: TypeScript/JavaScript, Python, Lua, Go, SQL, API design and OpenAPI, Microservices, Event-driven architecture, Serverless architecture, Linux system administration, Docker and CI/CD, Infrastructure as code

Systems, data and frontend: Unix/Linux and Bash, Networking, SQL indexes and stored procedures, Redis, React / Vue.js / Flutter, Testing, OAuth 2.0/OIDC/SAML/JWT, Physical security keys, Applied cryptography, Supply-chain security

Professional Experience

Senior Software Engineer — Encora Inc. (Consulting: enterprise e-commerce)

July 2023 – Present

- Own the product and user microservices for Alicorp, Peru's largest consumer goods company, supporting 2 backend tenants, 4 consumer applications, 35 distributors, ~3K products, ~150K users, and high-traffic core commerce services.
- Implemented user authentication, product search, navigation, pricing, and promotion workflows including combinable min/max price, combo, and payment-method rules.
- Implemented stored procedures and schema migrations; optimized hot queries across large relational datasets, improving average execution time by ~50%.
- Optimized under-tuned APIs and workflows, reducing peak-hour latency from up to 6s to sub-1s, lowering database CPU/RAM usage from ~85% to ~30% through indexing, Redis cache tuning with selective field loading, and Node.js Worker Threads.
- Improved production reliability with New Relic telemetry, structured logging, and targeted fixes, reducing incidents and maintaining near-zero rollbacks in owned repositories.
- Conducted technical interviews for senior engineering candidates, owning the technical evaluation process and delivering hiring recommendations to stakeholders.

Backend Developer — Blum SAF (Financial services)

January 2022 – June 2023

- Built event-driven financial microservices on AWS serverless infrastructure, covering API workflows, identity, storage, observability, secrets, and asynchronous processing.
- Designed API-first backend architecture and infrastructure as code with AWS CDK, converting product requirements into secure REST APIs, domain services, and repeatable cloud resources.
- Architected audit-sensitive financial workflows considering handlers, services, repositories, DTO validation, JWT authentication, secret isolation, and least-privilege IAM.
- Automated operational and developer workflows with TypeScript and Bash scripts, improving development and quality assurance time across local, development, testing and production environments.

- Delivered OpenAPI/Swagger documentation and Jest/Supertest test coverage across production integrations requiring reliability, auditability, and secure authentication.

Mobile Developer — INTECH (E-commerce)

February 2020 – December 2021

- Led Flutter and Dart mobile development for Android and iOS e-commerce applications, owning implementation from UI flows through release packaging.
- Migrated native Android Java functionality to cross-platform Flutter, reducing duplicated mobile code while preserving platform release requirements.
- Managed compilation, signing, store builds, Google Play Store deployment, production issue triage, and Figma-to-mobile implementation with non-technical stakeholders.

Full Stack / AI Developer — Freelance, Open Source and Personal Projects

2019 – Present

- Delivered fullstack web, automation, and infrastructure projects across backend services, frontend interfaces, cloud integrations, and developer tooling.
- Built 15+ AI-assisted automations across developer tooling, document workflows, audio analysis, reporting, scheduling, and structured data processing, preserving human review for decisions and final outputs.
- Automated interview-feedback drafting with diarization and LLM analysis, reducing a task that could take ~2 hours to roughly 10 minutes while keeping manual verification.
- Developed coding-agent workflows with parallel task execution, codebase search, context assembly, prompt rewriting, external APIs, scheduled automation, and persistent project memory.
- Designed RAG pipelines over codebases and documentation, extracting context from documents and images into structured prompts for actionable LLM outputs.
- Built AI-based ETL pipelines for extracting, normalizing, and transforming semi-structured content into reviewable outputs and downstream automation inputs.
- Ran direct model inference and benchmarking experiments on GPU-backed environments, comparing output quality, latency, and practical fit for automation workflows.
- Created automated document-generation workflows with LuaLaTeX, focusing on repeatable PDF builds and clean text extraction.
- Prototyped AWS-native agent workflows using Bedrock, Step Functions, Mountpoint for S3, IPv6-only networking, inference observability, and automated provisioning via Systems Manager.
- Secured TypeScript and Python dependencies with strict pinning, pnpm minimum-release-age, and uv exclude-newer to reduce exposure to newly published malicious packages.

Education

Systems Engineering — Bachelor's degree

San Cristóbal of Huamanga University

Languages

Spanish: native

English: proficient

Links

website: gedy.dev

linkedin: linkedin.com/in/gedy-palomino

github: github.com/gedzppelin