Ariel Magalso

AI Engineer

Negros Oriental, Philippines (Remote) | +63 915 616 9445 | hello@arielmagalso.com
arielmagalso.com | linkedin.com/in/magalsoariel | github.com/ArielMagalsoDev

6+ years software engineering experience
Python, FastAPI, PostgreSQL, Docker, Claude API, RAG, pgvector, Next.js, TypeScript, Evals

SUMMARY

AI engineer who ships production agentic systems - lead qualification, invoice controls, document-grounded Q&A - in Python and TypeScript on the Claude API, with RAG retrieval, schema-validated structured outputs, groundedness scoring, and prompt-injection defenses. Every system ships with a public evaluation suite and a human-in-the-loop on consequential decisions; six years of prior software engineering covers the range from data model to interface.

SKILLS

AI Systems: Claude API, OpenAI API, Claude Code, Codex, Retrieval-Augmented Generation (RAG), Vector search, pgvector, Structured outputs, Prompt engineering, Prompt caching, Evaluation suites, Groundedness and entailment scoring, Prompt-injection defenses, Human-in-the-loop workflows, n8n, Zapier

Application Engineering: Python, FastAPI, SQLAlchemy, Jinja2, TypeScript, JavaScript, Next.js, React, Tailwind CSS, Node.js

Data & Infrastructure: PostgreSQL, pgvector, SQL, Supabase, Docker, Docker Compose, Durable job queues, Rate limiting and spend caps, Per-stage audit logging, Document extraction

Integrations: REST APIs, Webhooks, OAuth, Accounting-system adapters, HubSpot API, GoHighLevel, Slack API, CRM integrations, Authentication and secrets management

Cloud & Delivery: Docker, Docker Compose, Linux VPS deployment and administration, Vercel, Cloudflare, DNS configuration, Git, GitHub

Practices: Evaluation-driven development, Failure-case analysis, Cost and latency optimization, Technical documentation

WORK EXPERIENCE

AI Engineer 2025 - Present
Self-employed - Negros Oriental, Philippines (Remote)

- Ship production AI systems end-to-end on the Claude API - Verdict (lead qualification) and LedgerGuard (invoice exceptions) in Python and FastAPI on PostgreSQL, containerized with Docker on a self-managed VPS, and Provenance (RAG with citations) in TypeScript on Next.js - each live with its own public eval suite and source code.

- Design decision pipelines around explicit evidence gates, escalating consequence by system: recoverable support answers are discarded below a groundedness floor, CRM-persisting scores are withheld below an evidence floor, and irreversible finance work carries no payment authority at all - people approve.
- Publish evaluation suites against labelled test sets - pass rates, per-decision cost, and documented failure cases - so accuracy and unit economics are verifiable rather than asserted.
- Engineer shared reliability layers: atomic-claim entailment scoring that discards output below a groundedness floor, schema-validated structured outputs, prompt-injection screening before generation, per-stage audit events, and rate limiting plus spend caps ahead of model calls.
- Integrate model output with systems of record over REST APIs - accounting-system adapters, document extraction, and webhook-driven handoffs - so decisions land where the business already works, and enforce that risky systems hold no write authority of their own.
- Own the full lifecycle: scoping, development, CLI tooling for ingestion and pipeline runs, containerized deployment to a self-managed VPS and to Vercel, DNS and subdomain configuration, evaluation, and documentation.

AI Automation Specialist / Web Developer 2024 - 2025

Self-employed - Negros Oriental, Philippines (Remote)

- Built a retrieval-augmented support chatbot over a 90+ passage knowledge base that answers only from verified sources with inline citations and escalates to a human when confidence is insufficient; handled 90+ real visitor conversations on a live site.
- Designed an n8n speed-to-lead automation that scores inbound enquiries 0-100, validates calendar availability in code rather than through the model, and triages leads hot/warm/junk - cutting response time from hours to under one minute.
- Built a Notion-to-web publishing pipeline that generates SEO metadata via the Claude API and writes status back to the source record - reducing publishing effort from 15-30 minutes to roughly 2 minutes per post.
- Integrated business platforms over REST APIs and webhooks, and documented each automation so non-technical owners could operate and monitor it.

Web Developer 2020 - 2024

Self-employed - Negros Oriental, Philippines (Remote)

- Built and shipped production web applications and client sites end-to-end, working across data models, back-end logic, and responsive front-end interfaces.
- Integrated third-party APIs, webhooks, and services to extend platform functionality, debugging live production systems to root cause.
- Authored technical and user-facing documentation for the systems delivered, reducing support requests and improving client self-sufficiency.

PROJECTS

Verdict - Lead Qualification (Python, FastAPI, Jinja2, SQLAlchemy, PostgreSQL, Docker Compose, Claude API): eleven-stage pipeline - identity resolution before model spend, fact extraction into a fixed schema, two-pass verification, an evidence-sufficiency gate, then deterministic points-based scoring. The model classifies; plain code does the arithmetic. Durable job queue built on a Postgres table with attempts and backoff rather than an external service; rate limiting plus a race-safe daily spend cap using row-level locking. 95% accuracy (57/60) on the eval suite at ~$0.01 per lead. Live: verdict.arielmagalso.com | Source: github.com/ArielMagalsoDev/verdict

LedgerGuard - Invoice Exception Automation (Python, FastAPI, PostgreSQL, Docker, Claude API, QuickBooks-style sandbox adapter): AI extracts invoice fields with every value evidence-linked to source text, then decimal-safe deterministic code recalculates totals, matches purchase orders and receipts, and catches renamed-rescan duplicates by normalized supplier identity. The workflow holds no payment authority - people approve, and bank changes require out-of-band verification. 35/50 held-out cases passed with 0% critical false clearances. Live: ledgerguard.arielmagalso.com | Source: github.com/ArielMagalsoDev/ledgerguard

Provenance - RAG Claim Verification (Next.js, TypeScript, Claude API, Postgres, pgvector): retrieves against a 52-passage approved policy corpus, decomposes each generated answer into atomic claims, and scores entailment against the retrieved evidence - below a 0.70 groundedness floor the answer is discarded, not softened. Eight inspectable stages route every request to answer, human review, or block, with prompt-injection attempts stopped before generation and an audit event persisted per stage. 45/45 committed eval cases, 100% route accuracy. Live: provenance.arielmagalso.com | Source: github.com/ArielMagalsoDev/provenance

Upskill Course - Online Learning Platform (Next.js, React, TypeScript, Tailwind CSS, Supabase, Claude API): course catalog, lesson delivery, and certificate logic with a citation-backed RAG assistant and automated content pipeline. Live: upskill-course-online.vercel.app

Cawyan Resort - Enquiry Qualification Automation (n8n, Claude API, REST APIs, webhooks): scores and triages inbound booking enquiries, validates availability in code, and dispatches replies - response time cut from hours to under one minute. Live: cawyanresort.com

EDUCATION

BS in Information Technology 2015 - 2019
Negros Oriental State University - Negros Oriental, Philippines