DON MARTIREZ — AI Engineer| GenAI Application & Systems Architect

Duarte, CA 91010 | don.martirez@gmail.com | (575) 459-1010 | Portfolio: don-the-imaginator.com/work/disney | artstache.app | the3pmdaily.com | linkedin.com/in/donmartirez

PROFESSIONAL SUMMARY

Full-Stack Software Engineer and AI Systems Architect with 10+ years of experience building web applications, backend microservices, and conversational AI platforms. **Built and shipped two production conversational AI platforms** — a GraphRAG learning assistant and a data-heavy financial intelligence platform with automated ingestion pipelines.

CORE SKILLS

- **GenAI & Agentic Engineering**: Python, FastAPI, Django, Flask, Prompt Logic, Structured Outputs, Conversational AI, Agentic Workflows, Function Calling, Model Context Protocol (MCP) Integration Patterns, LLM APIs (Claude, OpenAI, Gemini)
- **Retrieval & Grounding:** RAG, GraphRAG, Source-Backed Grounding, Embeddings, Semantic Chunking, Metadata Handling, Vector Search (MongoDB Atlas), Neo4j Knowledge Graphs, SQL/NoSQL, NL-to-SQL
- **LLMOps & System Reliability**: Evaluation Pipelines, Golden Sets, LLM-as-a-Judge, Regression Testing, Observability, Telemetry Logging, Latency & Cost Tracking, Production Troubleshooting, Gap Intelligence Telemetry
- **Cloud, DevOps & Tools:** Docker, Containerization (Cloud-Agnostic / Azure-Ready), Google Cloud Run, Azure Architecture Concepts, RESTful APIs, CI/CD, Git, Cursor, Claude Code
- **Frontend & Enterprise Standards:** React, Next.js, TypeScript, Tailwind CSS, WCAG 2.1 Accessibility, Enterprise Security Practices, Cross-Functional Technical Collaboration
- **Design**: High-Fidelity Prototyping, Adobe Creative Suite, Design Systems, Human-Centered Design

PROFESSIONAL EXPERIENCE

Founder & Lead AI Engineer — Retrieval & Grounding Frameworks

Artstache (artstache.app) | Remote | Jan 2024 – Present

- Platform launched Jan 2026 after 18-month R&D/build
- Built and maintained Python-based GenAI application components (FastAPI, LangGraph) delivering source-backed, context-aware responses via 5-node multi-agent DAG and Neo4j GraphRAG retrieval pipeline.
- Engineered document ingestion and semantic chunking pipelines with metadata handling, prompt logic, structured outputs, and strict validation patterns to enforce grounding.
- Implemented LLMOps practices including automated regression testing via LLM-as-a-judge evaluation framework with curated 50-prompt golden dataset (101 assertions in CI/CD), latency/cost logging, and production-ready Model Context Protocol (MCP) tool bindings.
- Connected conversational workflows to secure backend APIs and developed full-stack interfaces in Next.js, React, Tailwind CSS; use Cursor and Claude Code as required engineering workflow for velocity, testing, and documentation.

Founder & Systems Architect — Multi-Agent Data Processing Platform

The 3PM Daily (the3pmdaily.com) | Remote | Jan 2024 – Present

- Platform launched Jan 2026 after 18-month R&D/build
- Containerized Python multi-agent microservices using Docker to ensure seamless local-to-cloud environment parity and automated container deployment.
- Engineered automated data ingestion pipelines in FastAPI to ingest, synthesize, and structure complex operational and market data into user-ready financial intelligence.
- Built NL-to-SQL conversational interface over 45 tables & relational views, with curated schema catalog (schema_catalog.json) enriched with PascalCase JSON extraction and pre-aggregated views (v_income_statement_annual, v_balance_sheet_annual)
- Implemented dual self-healing repair: AST parsing + MySQL execution error detection triggers automatic 1-turn repair loop (fix_sql_error) passing error trace to Gemini 2.5 Flash for self-correction before rejection
- Developed Gap Intelligence telemetry (gap_logger.py) evaluating zero-row queries via LLM classifier to differentiate expected empty market filters from missing asset classes (e.g., Crypto, ESG), writing structured logs to data_gaps.json for observability
- Deployed cloud infrastructure on Google Cloud Run, managing environment configurations, REST API endpoints, and low-latency data persistence

Web Producer | The Huntington Library — San Marino, CA | Feb 2022 – Oct 2022

- Worked directly with stakeholders to understand business needs and gather feedback for major site redesigns — owned user experience including workflows, interfaces, automation, and WCAG compliance.
- Primary liaison between communications and engineering teams; documented technical decisions and tradeoffs clearly; experience across media and operations business areas.

Technology Coordinator | Anderson/LA (RR Donnelley)

Clients: Disney D23, Chevron, Wendy's | Jul 2015 – Dec 2015

- Delivered enterprise technology solutions for Disney D23 to optimize member database workflows and improve web-to-print production efficiency.
- Programmed relational database pipelines (SQL) and XML transformation scripts to handle high-volume member data adhering to enterprise security standards.
- Communicated technical choices and architecture trade-offs clearly to cross-functional stakeholders across media, retail, and operations.

PORTFOLIO & LIVE PRODUCTS

- **Artstache.app** (Watch Architecture & Agent Demo on LinkedIn): Generative AI educational platform using LangGraph + Graph RAG, golden sets, regression checks, LLM-as-judge.
- **the3pmdaily.com** (Watch Multi-Agent System Demo on LinkedIn): Multi-agent financial news platform — Python, Docker, FastAPI, Next.js, Google Cloud Run, RAG, agents, production eval pipelines.
- **don-the-imaginator.com** (Watch Interactive AI Portfolio Demo on LinkedIn): Interactive AI portfolio featuring integrated AI chat agent with custom tool-calling that renders dynamic,

generative UI components in real time — built with Next.js, LLM APIs, Cursor/Claude Code. Includes Loom demos.

- **thegoodgnome.com**: Full-stack e-commerce (Next.js, Stripe) built without dedicated engineering team — founder / builder proof.

EDUCATION

B.A. Media Arts & Design (Human-Centered Design Systems) — Mt. Sierra College, Monrovia, CA

CERTIFICATIONS

- Microsoft Certified: Azure Fundamentals (AZ-900) — In Progress (Exam Scheduled: October 2026)
- Snowflake Data Engineering Professional Certificate — In Progress (Target: Q4 2026)
- Google UX Design Professional Certificate — User Research, Wireframing, Figma High-Fidelity Prototyping, Usability Testing
- Google AI Professional Certificate — AI-driven data analysis, content creation, vibe coding custom AI solutions
- Generative AI Leader Certificate, Google Cloud — Strategic implementation of Gen AI agents, evaluation methods