DOMENICO DANIELE CELSA

GenAI Engineer | Multi-Agents Systems Architect

Milan, Italy | [+39 3282606575] | danielecelsa@gmail.com

LinkedIn | GitHub | My Portfolio

PROFESSIONAL SUMMARY

Senior Systems Engineer & Architect with 10+ years of experience in Cloud Infrastructure, Distributed Systems, and Technical Leadership. I am currently applying my background in high-availability systems to **GenAI Engineering**, specializing in **Autonomous Multi-Agent Systems** and **Agentic RAG**.

I combine a strong foundation in DevOps and Enterprise Integration (ex-Accenture Technical Architect) with modern AI frameworks (LangGraph, MCP) to build production-grade AI solutions that are resilient, observable, and cost-effective (**FinOps**).

CORE TECHNICAL SKILLS

- **GenAI Orchestration:** LangGraph, LangChain, Model Context Protocol (MCP), ReAct Agents, Supervisor-Worker Patterns.
- **Retrieval & Data:** Agentic RAG, Hybrid Search (BM25 + Vector), Cross-Encoder Reranking (FlashRank), ChromaDB, SQLite (Schema Introspection).
- **Backend & Engineering:** Python (AsyncIO), FastAPI, Pydantic, Server-Sent Events (SSE), SQL, Docker, CI/CD.
- **Observability & FinOps:** Custom Callback Handlers, Distributed Logging (Redis/BetterStack), Token-Cost Attribution, Unit Economics Analysis.
- **Infrastructure:** AWS (EC2/S3/Cloud Architecture), Linux/Unix, TCP/IP Networking.

ENGINEERING CASE STUDIES (GEN AI)

*Detailed architecture and code available at: **danielecelsa.github.io/portfolio***

Hierarchical Multi-Agent Orchestrator | *LangGraph, SQLite, Gemini*

- **Architecture:** Designed a **Supervisor-Worker** state machine to decompose ambiguous natural language requests into executable actions across Calendar, Email, and SQL agents.
- **Reliability:** Implemented **deterministic SQL Schema Introspection**, allowing the agent to write syntactically correct queries and resolve entities without hallucinations.
- **FinOps:** Engineered a custom observability middleware to track **Token Usage & USD Cost** at a per-agent level, enabling precise ROI calculation for every interaction.
- Try the App

Decoupled Autonomous Research Agent | *Model Context Protocol (MCP), FastMCP*

- **Microservices Design:** Built a client-server architecture strictly following the **MCP Standard**, decoupling the LLM reasoning engine (Client) from the tool execution environment (Server).
- **Hybrid Grounding:** Developed logic to cross-reference AI-generated summaries against raw search metadata (URLs/Snippets), ensuring 100% citation accuracy and eliminating fabrication.
- **Protocol:** Utilized Server-Sent Events (SSE) for real-time protocol tracing, providing full visibility into the agent's thought process over HTTP.
- [Try the App](#)

Enterprise Document Intelligence (Agentic RAG) | *ChromaDB, LangChain*

- **Privacy-First:** Architected ephemeral, **session-scoped Vector Stores** to ensure strict data isolation between concurrent users.
- **Precision:** Implemented a **Recursive Retrieval** loop combining Hybrid Search and Cross-Encoder Reranking to extract high-precision context ("needle in a haystack") from long documents.
- **Cost Optimization:** Designed an autonomous decision loop where the agent evaluates *necessity* before retrieval, reducing unnecessary token consumption.
- [Try the App](#)

Multimodal Compliance Auditor *(In Development)*

- Engineering an AI Auditor capable of native video/audio ingestion for automated risk assessment on long-context multimedia. Enforces strict JSON schemas via Pydantic for automated downstream compliance reporting.

PROFESSIONAL EXPERIENCE

Coderblock (Metaverse Startup) | *Customer Success Manager & Community Lead*
Palermo, Italy | 01/2023 – 10/2024

- **Technical Bridge:** Translated complex architectural constraints into business value propositions, serving as the primary liaison between Engineering and Enterprise clients.
- **Process Automation:** Engineered internal workflow automations (Zapier/Webhooks) to synchronize data pipelines between Marketing and Sales, significantly reducing manual overhead.
- **Growth Engineering:** Architected community infrastructure and executed multi-channel growth strategies, converting community members into high-value investors ("Land Owners").

The Sports Club (Web3) | *Web3 Operations Lead & Technical Community Manager*
Milan, Italy | 02/2021 – 02/2022

- **Distributed Systems:** Managed end-to-end technical setup for NFT launches on **Solana**, including Candy Machine configuration and JSON Metadata generation.
- **Community Architecture:** Architected secure Discord infrastructure (permissions, security bots, ticket systems) to support and entertain thousands of concurrent users.
- **Tokenomics:** Collaborated on token economy design and utility structures for the ecosystem.

Accenture Consulting | *Infrastructure Consultant / Technical Architect*
Milan, Italy | 06/2015 – 02/2020

- **Mission-Critical Migration:** Led the end-to-end migration of legacy on-premises infrastructure to Aruba Cloud for a major enterprise client, ensuring **Zero-Downtime** cutover strategies for high-availability production systems.
- **Technical Leadership:** Acted as **Single Point of Contact (SPOC)** leading a squad of 7-8 engineers. Translated complex business requirements into technical execution plans for global TelCo providers.
- **DevOps & Reliability:** Designed scalable AWS environments and optimized CI/CD workflows, significantly reducing deployment friction. performed root cause analysis and advanced troubleshooting for distributed systems.

EDUCATION

Master's Degree in Telecommunication Engineering (110/110)
Università degli Studi di Palermo (2011 - 2015)

- **Thesis:** Vehicular Ad-Hoc Networks (VANET) – Researched decentralized mesh networking protocols and latency optimization in mobile environments.

Bachelor Degree in Telecommunication Engineering
Università degli Studi di Palermo (2007 - 2011)

LANGUAGES

- Italian (Native), English (Professional Working Proficiency), Spanish (Colloquial).