Manuel Scionti

Senior AI Engineer | Barcelona, ES (EN C1 | ES C1 | IT native) | open to remote EU
scionti.manuel@hotmail.it | linkedin | mnlscn.github.io | github

Summary

AI Engineer shipping production GenAI features users actually feel: retrieval, agent workflows, evaluation, and the production hardening that keeps LLM systems reliable, cost-efficient, and compliant at scale. Currently owning the offline batch AI FAQ platform at ADP – **1000+ daily users** across US & EMEA under GDPR data-residency and zero-hallucination requirements.

Featured Project – ADP AI FAQ Platform

Offline batch RAG pipeline generating pre-validated FAQ content from 15+ business units' docs (payroll, HR, compliance), served via embedded chatbots, smart searchbar, and an internal API. **Hybrid OpenSearch retrieval** (keyword + vector), **dedicated reranker**, **LLM-as-judge eval gating** offline before any change ships. In production for 6 BUs, **1000+ daily users** US & EMEA, **+20%** retrieval quality on the golden set, **-50%** integration time for new BUs, **-36%** CI/CD build time. *Key bets:* in-house orchestration framework over LangChain (full control on evals & observability), reranker over a frontier model (same lift, fraction of the cost), hybrid over pure semantic on technical / policy queries. Team of 5 – owned eval suite, retrieval / reranking, CI/CD modernisation.

Core Skills

LLM & Agents: Production RAG, Graph RAG, agent orchestration, multi-agent workflows, LLM-as-judge, eval pipelines & golden datasets, prompt engineering, semantic / hybrid search, rerankers, model routing, prompt & cache cost optimisation
Backend: Python, FastAPI, Pydantic, REST / serverless APIs, SQL, Hydra, YAML config
Cloud & MLOps: AWS (Lambda, API Gateway, SageMaker, Bedrock), Azure, Databricks, OpenSearch, Jenkins / Bitbucket CI-CD, observability, A/B testing & evaluation gating
ML & Multimodal: NLP, computer vision, speech (TTS / STT), multimodal retrieval, clustering

Selected Achievements

- **IBM × Datapizza GenAI Hackathon (Milan, 2025)** – top 10% of 600+ applicants, Graph RAG project – demo
- **Open-source contributor, IBM/Agentics** – **3 PRs merged into main** (production bug in `PydanticTransducerVLLM`, dependency cleanup, additional fixes) – repo
- **Tech lead & organiser, Hackatania GenAI Hackathons (2 editions)** – ed. 1 | ed. 2

Professional Experience

AI Engineer Mar 2025 – Present
ADP – Automatic Data Processing Inc. (via Capitole Consulting) Barcelona, Spain

- **Owned end-to-end** the offline batch AI FAQ platform across 15+ BUs; **6 BUs in production**, **1000+ daily users** US & EMEA under GDPR data-residency and zero-hallucination requirements
- **Core contributor** to the in-house LLM orchestration framework (chosen over LangChain for full control on evals, observability, provider swapping); reusable abstractions **cut new pipeline integration time by 50%**
- **Designed** the LLM-as-judge eval suite (10+ evals, golden dataset) gating every prompt / retrieval change; lifted retrieval quality **+20%** via hybrid OpenSearch (keyword + vector) and a dedicated reranker

- **Refactored** Jenkins CI-CD into modular stages (unit, integration, security, packaging, Artifactory, docs); build time **25 → 16 min (-36%)**
- Led an internal workshop on AI coding agents (20 attendees); shipped shared workflows for doc-to-code sync

AI Engineer Nov 2023 – Mar 2025
Neodata Group Catania, Italy

- **Led the design of Neovid**, a multimodal retrieval system over image, video, and document collections (NLP, CV, face recognition, TTS, STT); processed **1000+ videos** for Italy's largest broadcaster
- **Built an internal extraction framework** for tables and enterprise documents; **cut table-extraction errors by 75%** vs OSS baselines (pdfplumber, PyMuPDF) and outperformed Azure Document Intelligence on structured specification tables
- **Main developer of NeoKnowledge**, an AI pipeline generating personalised educational content (text, audio, video) deployed across **5+ museums**
- **Led 2 junior engineers** from pilot to production, owning the AI architecture and shipping **4 projects** end-to-end

Machine Learning Engineer Feb 2023 – Nov 2023
Koexai Catania, Italy

- Built end-to-end ML pipelines on AWS (S3, Glue, Athena) for the largest coworking space in Catania (€1M revenue); developed segmentation models (K-Means, Hierarchical, DBSCAN) over 500+ customers
- Automated recurring data-entry workflows, **cutting operational time by 80%**

Education

M.Sc. Data Science, University of Catania – 110/110 *summa cum laude* (GPA 4.0) 2021 – 2023
Erasmus+ in AI, Goethe University Frankfurt 2023
B.Sc. Business Management, University of Catania – 100/110 (GPA 3.3) 2016 – 2020

Certifications

LangGraph (LangChain, 2024) | LangChain in Action (2024) | Intermediate SQL (2023) | Pre-Security, TryHackMe (2025)