

DAVID FARCAS

AI ENGINEER

Cluj-Napoca, Romania | +40 720 806 333 | farcas.david@yahoo.com | linkedin.com/in/farcas-david | davidfarcas.ro

EXPERIENCE

AI Engineer | RebelDot Nov. 2024 – Present

Cluj-Napoca, Romania

RAG-based Email Response Automation Aug. 2026 – Present

Automotive marketplace client | Python, DSPy, AWS Bedrock, PostgreSQL/pgvector

- Independently architected and delivered an **11-category RAG email assistant**, combining classification, pgvector/Cohere retrieval and Haiku reranking to generate cited replies for human review.
- Raised **correct drafting from 45% to 88%** and **top-5 retrieval from 70% to 91%** by turning historical agent replies into a reusable, source-grounded knowledge base.
- Replaced ad hoc prompt tuning with **DSPy/GEPA programmatic optimization** against defined classification and safety metrics, making prompt selection repeatable and quality changes measurable.
- Added **GLiNER/Presidio PII masking**, **MPNet clustering** and **DeBERTa NLI** before indexing agent replies, limiting personal-data exposure and keeping knowledge-base answers traceable to source evidence.

LLM-based Call Quality Assurance 2026

Insurance client | Python, LangChain, DSPy, HuggingFace Transformers

- Built a **rubric-driven LLM scoring system** that turns call transcripts into weighted, per-criterion assessments with cited evidence, standardizing quality reviews and making scores auditable.
- Designed **RapidFuzz and DeBERTa NLI verification** to flag fabricated citations and unsupported reasoning **without extra LLM calls**; local NLI checks benchmarked at **108–151 ms per criterion**.
- Implemented **DSPy/GEPA prompt optimization** against macro-F1 and Cohen’s kappa; benchmarked **eight criteria across 119 calls (952 judgments)** to track human agreement and identify quality regressions.

AI Applications & Platform Engineering Nov. 2024 – Dec. 2025

Insurance, automotive & enterprise clients | Python, FastAPI, Streamlit, AWS/Azure, Terraform

- Delivered **19 self-scoped optimization tickets**, cutting **2.5M-row queries from 25s to under 2s**; expanded filtering from **3 to 40+ parameters** under a Q4 deadline using PostgreSQL, batched APIs and Redis.
- Delivered **multimodal OCR and cloud AI applications** using FastAPI, Streamlit, Terraform and Azure Pipelines; added Microsoft Entra ID/JWT authentication, LLM caching and cross-region AWS Bedrock load balancing.

Engineering & collaboration: Established shared AI coding-agent rules and pre-commit checks for style, tests and commits; wrote architecture and contribution guides adopted by the team.

SKILLS

Python RAG LangChain LangGraph
DSPy/GEPA AWS Bedrock pgvector

LLM engineering: Prompt optimization, structured outputs, grounding, reranking, evaluation; LiteLLM, OpenWebUI.

Models & NLP: Claude Sonnet/Haiku/Opus; Cohere Embed English v3, MPNet, DeBERTa NLI, GLiNER, Presidio, spaCy, PyTorch, Sentence Transformers, HuggingFace Transformers, BERTopic, UMAP.

Backend & data: SQL, FastAPI, Pydantic, Streamlit, SQLAlchemy, PostgreSQL, Redis, Qdrant, Pinecone, FAISS, pandas, NumPy, scikit-learn, SciPy, RapidFuzz, BM25, ROUGE-L.

Cloud & delivery: AWS: S3, RDS, EC2, Secrets Manager. Azure: App Service, ACR, Key Vault, Entra. Terraform, Docker, Podman, Kubernetes, Azure DevOps, Azure Pipelines, Git, Bash, pytest, Ruff, mypy.

PROFESSIONAL STRENGTHS

End-to-end ownership; initiative; analytical problem-solving; technical communication; delivery under deadlines.

Languages: Romanian, English.

CERTIFICATIONS

Claude Certified Architect - Foundations

20 Jun 2026 – Jun 2027

Agentic orchestration with Claude API/Agent SDK; tools and **MCP integration**; Claude Code workflows; prompt engineering, structured outputs, context management and reliability.

Microsoft Certified: Azure AI Engineer Associate (AI-102)

Earned 2024

Azure AI design and deployment; NLP, computer vision, document extraction, search and responsible AI.

EDUCATION

Technical University of Cluj-Napoca

2020 – 2024

Bachelor’s Degree in Electronics, Telecommunications and Information Technology.