Ioan-Vlad Pantiru

AI Engineer | Cluj-Napoca, Romania | Open to EU remote
+40 752 532 048 | ioanvladpantiru@gmail.com
linkedin.com/in/ioan-vlad-pantiru-82b2361a1 | github.com/ioan-vlad-pantiru

AI engineer shipping production LLM systems (LangGraph agents, RAG, Azure OpenAI) end to end, from data pipelines and verification gates to full-stack apps and CI/CD. Former technical co-founder of an edge-AI startup.

Experience

AI Solution Engineer Jul 2025 – Present
Computacenter (joined as AI Intern, Jul – Aug 2025) *Cluj-Napoca*

- Lead developer of an **AI-powered IT asset management platform** (Next.js, FastAPI, LangGraph, Azure OpenAI, PostgreSQL) that turns 20K+ device client inventories into specs, refresh plans and ROI reports; shipped to production in ∼**2 months**, used by ∼15 consultants, identifying ∼**$600K in savings per client**.
- Cut per-client device research from ∼**2 days to ∼1 hour** with LangGraph pipelines combining vendor APIs, web-grounded search (Grounding with Bing) and LLM spec extraction, resolving launch dates for ∼**80%** of models; a confidence gate returns "TBC" instead of unverified specs.
- Built an LLM-assisted importer for enterprise inventory/CMDB workbooks: exact, then fuzzy, then AI-proposed column mappings with human confirmation, plus cleaning of garbled, float-corrupted and misspelled values.
- Added LLM task routing and telemetry across 13 AI modules, Celery/Redis batch enrichment, Entra ID SSO, and ∼**1,300 automated tests**; cut CI costs by moving GitHub Actions to self-hosted ARM64 runners.
- Delivered Azure AI Foundry solutions: a RAG recruitment assistant (pgvector, Redis) integrated with SharePoint and SuccessFactors, a ServiceNow AI Copilot, and a multi-agent device-data enrichment system.

Freelance Full-Stack & AI Engineer Jun 2026 – Present
Self-employed *Remote*

- **MitMed clinic platform** (mitmed.ro): sole developer of a client portal and back office (Next.js, FastAPI, PostgreSQL), live in ∼**3 weeks**, serving **60–70 clients/month**; made the clinic fully paperless (booking with Google Calendar sync and WhatsApp reminders, GDPR consent management, PayU billing), with 130+ tests and CI/CD auto-deploy.
- **Construction management platform** (in development, FastAPI): end-to-end system for a mid-size construction company (stock, crews, projects, cost control) with LLM document extraction, grounded bid drafting where every figure cites a company record, and e-Factura/REGES integrations.

Technical Co-Founder Mar 2025 – Feb 2026
Horizon, AI heads-up display for motorcyclists *Cluj-Napoca*

- Built a fully offline perception pipeline on a Raspberry Pi 4: YOLOv8s trained on 70K BDD100K images and quantized to INT8 TFLite, SORT tracking, ego-motion compensation and a risk engine ranking hazards on a helmet HUD, reaching **6 FPS** effective throughput.
- Took the product from concept to working hardware prototype (basis of my bachelor's thesis); **finalist at Innovation Labs 2025 and Stage Two (Berlin)**. Paused after a market analysis showed rising hardware costs pushed the price beyond what riders would pay.

Projects

HubShift | ***1st place**, GreenPow challenge, 28DIGITAL (EIT Digital) kick-off, Tallinn* Oct 2026

- Team Watt Matters (6): GPU pod running AI batch jobs on the ∼94% idle grid capacity of EV fast-charging hubs, yielding to cars within seconds; delivered pitch, business model, live operator dashboard and pilot plan.

WalkSafe | *React Native, Django, PostgreSQL, OpenAI, scikit-learn, osmnx* Oct 2024 – Feb 2025

- Safety- and air-quality-aware pedestrian routing app; **Special Mention**, M100-Cities Hackathon.

Skills

AI/LLM: LangGraph, LangChain, Azure OpenAI, Azure AI Foundry, RAG, multi-agent systems, pgvector, ChromaDB
ML/CV: YOLOv8, TensorFlow Lite (INT8 quantization), MobileNetV3, OpenCV, Kalman/SORT tracking
Programming: Python, TypeScript, JavaScript, SQL, Java, C++
Web & Backend: FastAPI, Next.js, React, React Native, PostgreSQL, SQLAlchemy/Alembic, Celery, Redis
Cloud & DevOps: Azure (Container Apps, Entra ID, Key Vault, App Insights), Docker, GitHub Actions, pytest
Languages: Romanian (native), English (C1 Advanced, Cambridge English Scale 199)

Education

EIT Digital Master School, MSc in Cybersecurity (double degree) Oct 2026 – 2028 (expected)
Babes-Bolyai University (year 1), University of Twente (year 2) *Cluj-Napoca / Enschede*
Babes-Bolyai University Oct 2023 – Jul 2026
BSc in Artificial Intelligence (taught in English) *Cluj-Napoca*

- Thesis: *Horizon: A Real-Time Multi-Stage Perception System for Motorcycle Rider Safety Using Edge AI.* Novel lateral-closing-speed and erratic-behaviour risk scores; 306 tests run on the Pi via self-hosted CI.