

Pierre-Jean Martin

AI / LLM Engineer

Production LLM infrastructure: multi-agent pipelines, RAG, and self-hosted LLMOps.

pierrejean.remy.martin@gmail.com | +41 76 651 68 62 | linkedin.com/in/pierre-jean-martin | Lausanne, Switzerland
French citizen (EU).

TECHNICAL SKILLS

AI / LLM

LangGraph (multi-agent, deterministic pipelines), RAG, hybrid search (pgvector), prompt engineering, LiteLLM (multi-provider routing, virtual keys, budgets), LangSmith tracing, MCP. Frontier (OpenAI, Anthropic, Gemini) and open-weight (MiniMax, GLM, Kimi) models. Image generation.

Data Platform

Palantir Foundry: Pipeline Builder, Ontology Manager, AIP Logic, OSDK, Contour, Automation.

Backend / Full-stack

Python, FastAPI, Next.js, REST APIs, n8n.

Infra / DevOps

Docker, Docker Compose, Cloudflare Tunnel + Access, nginx, CI/CD, GHCR, self-hosted deployment.

Also

C, C++, Java, JavaScript, MATLAB, Git. Computer vision (MediaPipe, Bi-RefNet).

CERTIFICATIONS

- Palantir Foundry Data Engineer, Professional (Certified)
- Claude Certified Architect (in progress)

EDUCATION

EPFL, M.Sc. Robotics · 2024

Mechatronics, Robotics & Automation

EPFL, Minor, Management, Technology & Entrepreneurship · 2024

EPFL, B.Sc. Microengineering · 2020

LANGUAGES

French · native

English · full professional

German · limited working

LEADERSHIP

President and board member of EPFL student associations (2016–2019), organizations running on 2M+ CHF annual budgets: led teams of up to 45 people and ran events of up to 30k CHF. Recipient of the EPFL President's award for exceptional involvement.

PROFILE

AI engineer specializing in production LLM systems: deterministic multi-agent pipelines (LangGraph), RAG and hybrid retrieval, and full LLMOps stacks (LiteLLM, multi-provider routing, tracing, per-team budgets). I ship end to end, from architecture to self-hosted deployment, with a bias toward data that stays in-house. I also design and deliver hands-on AI training that helps companies and their teams adopt these tools with confidence.

EXPERIENCE

The Original Body AG, Lausanne

Nov 2025 - May 2026

AI Tech Lead, LLM Platform & Engineering

Promoted from AI Systems Engineer (Nov–Dec 2025) to AI Tech Lead (Jan–May 2026)

- Led a team of up to **5 developers** across all AI, automation, and data initiatives, owning technical direction end to end.
- Designed and shipped **12 deterministic LangGraph pipelines** in production, including a ticket-drafting pipeline handling **1,000+ support tickets per month** and a content pipeline generating knowledge-base articles from a corpus of **~15,000 Zoom coaching transcripts**; delivered in dual-deploy mode (standalone server plus in-process package).
- Built a **self-hosted knowledge-management platform** connected to the company stack (Slack, Gmail, Drive, Palantir Foundry, Jira, Confluence), with streaming article generation, transcript summarizer, in-chat image generation, and AI user/team memory, all behind Cloudflare Access.
- Set up the full **LLMOps stack**: LiteLLM as a unified proxy (multi-provider routing, per-pipeline virtual keys, per-team budgets), LangSmith tracing, RAG and hybrid retrieval (pgvector), and n8n orchestration connected to agents via MCP.
- Packaged a computer-vision posture-analysis pipeline (MediaPipe Pose + Bi-RefNet) as FastAPI + Docker and deployed it as a Palantir Foundry transform, running **~500 analyses per month**.
- Managed self-hosted infrastructure (Docker Compose multi-stack, Cloudflare Tunnel + Access, nginx, CI/CD with fast rollback) and built interfaces letting non-technical teams iterate on AI pipeline behavior.

État de Vaud, Étoy

Jan 2025 - Feb 2026

Substitute Teacher, Gymnase

- Taught Mathematics, Computer Science, and Office tools across all levels (years 1–3), standard and advanced tracks; held responsibility for a class.

Private Tutor, Programming / AI / STEM

Sep 2016 - present

- 50+ students from EPFL, HEC, HES-SO, EHL and professionals reskilling; 100% pass rate, in French and English.
- Programming (C, C++, Python, Java) plus "Coding with AI": Claude Code, Copilot, Cursor, and prompt-engineering fundamentals, moving learners from AI users to builders.

EARLIER EXPERIENCE

- Master's Thesis**, Microcity Neuchâtel (2023–2024): design and control of a robotic orbital eye prosthesis.
- R&D Manager & Business Analyst**, Time Designers, Lausanne (2021): bistable actuator design; watch-market analysis; two EPFL startup accelerators.
- Automation Engineer**, Vigyan Ashram, India (2022).
- Teaching Assistant**, EPFL (2016–2021).