

Adrian Zaplata - AI Engineer

Poznań, Poland | +49 174 608 8033 | adrianzaplata@gmail.com

LinkedIn: [linkedin.com/in/adrianzaplata](https://www.linkedin.com/in/adrianzaplata) | GitHub: github.com/adrianzaplata | Portfolio: azaplata.dev

AI Engineer with 7+ years of production software engineering experience. I ship RAG pipelines, LLM agents, MCP servers, automations, and Claude Code integrations across web, mobile, 3D, and real-time systems.

Open to: Full-time, contract, freelance | Global remote

Languages: English (fluent), German (fluent), Polish (native)

TECHNICAL SKILLS

AI & LLM Integration: API integration, Claude Code, Instructor + Pydantic (structured extraction outputs), Prompt Engineering (rejection rules, prompt versioning, decomposition), MCP (Model Context Protocol), Ollama (local AI models)

Retrieval-Augmented Generation (RAG): vector databases, embeddings, pgvector, cross-encoder reranking (sentence-transformers), section-based chunking, RAGAS evaluation (4 metrics, golden dataset), semantic search

Agent Orchestration: LangGraph (ReAct agents, streaming, tool calling), LangChain, FastMCP (stdio server development)

Observability & Evaluation: Langfuse (LLM tracing), structlog, pip-audit

Backend & Full-Stack: Python (async, Pydantic, pydantic-settings, SQLAlchemy, asyncpg, tenacity), FastAPI (SSE via sse-starlette, auth, rate limiting), Alembic, Typer, TypeScript, React, React Native, Supabase, Docker, PostgreSQL, C#, C++

Cloud & DevOps: Azure (Container Apps with scale-to-zero, Azure Database for PostgreSQL Flexible Server + pgvector, Static Web Apps, Key Vault, Log Analytics, Entra External ID, Azure Blob state backend), Terraform (modular per-env composition, Azure Verified Modules, tfint, tfsec, two-pass CORS bootstrap), GitHub Actions (CI/CD, OIDC), GHCR

ML & Image Generation: PyTorch, Stable Diffusion 1.5, LoRA fine-tuning (kohya_ss), Hugging Face

Software Engineering: Git, GitHub CLI, pytest, Ruff, Pyright, uv, Playwright, Maestro, Software Architecture, Performance Optimization, Unity, Unreal Engine, 3D Graphics

PROFESSIONAL EXPERIENCE

AI Engineer | Freelance

March 2025 - present | Remote | Berlin, Germany

Developing RAG systems, LangGraph agents, MCP servers, Cloud deployments, FastAPI, prompt engineering, n8n automations, Claude Code integrations, AI knowledge sharing.

- Personal Interactive Agent - Chat Widget on my portfolio website, answering questions about my background - it's a Cloudflare Pages Function that calls Anthropic API with a system prompt based on my bio, hard refusal string for off-topic prompts, guardrails (Cloudflare Turnstile IP rate limit, daily USD spend cap in KV, token limit) | Website: azaplata.dev
- Job RAG - RAG System that ingests job postings, extracts structured data, compares the data to a profile, and exposes the corpus through a LangGraph agent, FastAPI service, and MCP server. Provisioned on Azure via Terraform with OIDC-federated CI/CD. Evaluated with RAGAS against a golden dataset | GitHub: github.com/adrianzaplata/job-rag
- Ploop - AI Game Generation Platform where you type a game idea and get a playable, mobile-friendly game at a shareable link in ~15s. Versioned prompts, A/B prompt evaluations, sandboxed iframe | Website: tiktok-game-eta.vercel.app
- Relay - A miniature real-time telemetry platform for a connected e-bike fleet. Built as a focused, end-to-end vertical slice of a greenfield, cloud-native micromobility platform. Events flow through Kafka, ingestor writes to PostgreSQL database, FastAPI serves a React/TypeScript operator console | github.com/AdrianZaplata/relay

Senior Software Engineer | Automotive HMI & ADAS | Siili Solutions

March 2022 - March 2026 | Remote | Berlin, Germany

Software development and 3D graphics for embedded systems - Automotive HMI and ADAS for a German premium automotive OEM. Built Claude Code pipeline with MCP integration automating feature implementation, compilation, and testing.

Senior Software Engineer | Freelance | Automotive HMI & ADAS | CEER

June 2025 - December 2025 | Remote | Riyadh, Saudi Arabia

Software development, 3D graphics and technical consultancy for embedded systems - Saudi Arabia's first electric vehicle brand.

Junior Software Engineer | VR & PC projects | NeXR Technologies SE

September 2020 - March 2022 | Hybrid | Berlin, Germany

Developed a virtual live concert and VR projects.

EDUCATION

AI Coding for Real Engineers - Matt Pocock (2026) - A two-week cohort for developers who want to use AI coding tools on real production-style work. Context gathering, planning, steering, feedback loops, AFK agents, and human-in-the-loop review.

Anthropic (2026) - Claude Code in Action

IBM / Coursera (2025) - Introduction to Data Engineering

Poznań University of Life Sciences (2012 - 2017) - B.Eng. Landscape Architecture