TUGAY SERINKEN

tugay.serinken@proton.me | LinkedIn | GitHub | Langs.: German, English, Turkish

Senior AI Engineer with a Staff-level fullstack background who takes AI products from **prototype to production** as a freelance technical lead across **11+ years**. Builds agentic workflows, RAG systems and fine-tuned LLMs on Azure, and runs them efficiently at scale — combining full product ownership with high-performance architecture to deliver measurable business outcomes.

KEY ACHIEVEMENTS

- Saved **1,200+ manual hours per month** with enterprise AI agents built on Copilot Studio and Azure OpenAI (ConferenceDirect)
- Automated support for **50,000+ daily queries** with an enterprise RAG system, cutting **resolution time by 65%** (ConferenceDirect)
- Reduced **inference latency by 45%** and **cloud GPU costs by €12K/month** with TensorRT and ONNX (ConferenceDirect)
- Built a **multi-agent LLM system** for SEO analysis in Seobase, a product driving **€1.1M yearly revenue** (Teamstack GmbH)
- Drove **€10M+ annual revenue** by building a low-code platform powering **1,000+ websites** (Compado GmbH)

SKILLS - *All professional proficiency or above.*

Languages: Python | TypeScript | Go | SQL | Rust | C++ | Bash
AI Frameworks: PyTorch | Hugging Face Transformers | LangChain | LlamaIndex | ONNX Runtime
Generative AI: Copilot Studio | Azure OpenAI | Fine-tuning (LoRA/QLoRA) | RAG | Agentic Workflows | MCP | Prompt Engineering
AI Infrastructure & Cloud: Azure (AI Foundry, AI Search, ML) | AWS Bedrock | Docker | Kubernetes | TensorRT | GitHub Actions
Databases: PostgreSQL | Pinecone | Weaviate | Cosmos DB | Redis | MongoDB
Fullstack: React | Next.js | Angular | Vue | Node.js | FastAPI | Django | GraphQL | REST | Microservices

WORK EXPERIENCE - *Over a decade of experience.*

Senior AI Engineer (Freelance part-time) at ConferenceDirect Germany GmbH | Frankfurt **Mar 2021 - Present**

- Architected and deployed enterprise AI agents with Microsoft Copilot Studio and Azure OpenAI on secured internal knowledge bases, automating HR and IT workflows and saving 1,200+ manual hours per month.
- Designed a RAG system on Azure AI Search and Llama 3 for the company's attendance platform (300K users, 13,000+ events a year), automating support for 50,000+ daily queries and cutting ticket resolution time by 65%.
- Fine-tuned Mistral LLMs via QLoRA on corporate data with Azure ML compute, improving domain-specific accuracy by 22%.
- Optimised inference with NVIDIA TensorRT and ONNX, cutting latency by 45% and cloud GPU costs by €12K per month.
- Mentored 6 developers, introduced agentic coding with MCP servers, and set up CI/CD for AI models with GitHub Actions.

Fullstack Engineer (Freelance) at Kapsys GmbH | Vienna **Jan 2025 - Jun 2026**

- Replaced an old enterprise monolithic CRM of an insurance company in Switzerland by creating Microservices, which is now used by over 120k insurance brokers for digitising the management of clients, insurance policies, and broker commissions.
- Automated commission calculation and renewal tracking, cutting administration time by 35%, on a stack of Angular and Python.

AI Engineer (Freelance) at Teamstack GmbH | Vienna **Sept 2023 - Jan 2025**

- Built an autonomous multi-agent system on early LLM APIs that scrapes public SEO data, runs analyses and generates automated summaries for the enterprise SEO platform Seobase (€1.1M yearly revenue).
- Integrated NLP pipelines into existing enterprise software, improving search relevance for internal documentation by 30%.
- Packaged and deployed 15+ microservices for AI inference with Docker and Kubernetes, maintaining 99.9% uptime.
- Led and mentored a team of 4 engineers from requirements to release, incl. Seobase's Angular-to-Next.js migration.

Lead Frontend Engineer (Freelance) at Interwetten Ltd. | Vienna **Sept 2023 - Mar 2024**

- Built a micro-frontend blog platform that renders fast multilingual, multi-licence pages from a headless CMS.
- Improved the TTFB by 25% and reached a mobile PageSpeed score of 94% with 100% in all other areas.

Senior Frontend Engineer (Freelance) at 1&1 Telecommunication SE | Karlsruhe **May 2023 - Sept 2023**

- Led the design-system POC on Stencil.js web components, enabling reuse across 5 products and cutting dev time by 40%.

Senior Frontend Engineer at Compado GmbH | Berlin **May 2021 - May 2023**

- Built a Wix-like low-code page editor (Vue.js, Stencil.js) powering 1,000+ comparison sites with €10M+ revenue in 2022.

Lead Fullstack Engineer (Freelance) at Cassis Media GmbH | Ratingen **Sept 2019 - May 2021**

- Modified the Snapify mobile app for influencers (36,000 downloads), reduced bug reports by 46%, improved load time by 32%.
- Built an admin panel web app from scratch using React, Redux, Go, and Docker.

Fullstack Developer at Incognito GmbH | Hannover **Jan 2015 - Sept 2019**

- Built an enterprise resource management tool from scratch (team of 4) now used by over 250 dialysis centres in Germany.
- Replaced a monolithic backend with a microservice architecture, reducing API response time by 60%.

EDUCATION - *Hands-on, interdisciplinary education.*

Bachelor of Science in Business Informatics at Hochschule Karlsruhe **2015 - 2017**
IHK: Fachinformatiker Anwendungsentwicklung at b.i.b. International College Hannover **2012 - 2015**