Berat Erkan Elcelik

AI Engineer / Data Scientist / Entrepreneur

Anne-Conway straße 17, Bremen 28359, Germany · +49 155 1019 8308 · elcelikberaterkan@gmail.com

Links

LinkedIn: https://www.linkedin.com/in/beraterkanelcelik

GitHub: https://github.com/beraterkanelcelik

Education

University of Bremen
M.Sc., Artificial Intelligence and Intelligent Systems
10/2024 – Present
Bremen, Germany

Ankara University
B.Sc., Computer Engineering (GPA: 3.76/4)
10/2018 – 08/2023
Ankara, Turkey

Profile

- Master's student in Artificial Intelligence with hands-on experience in ML, reinforcement learning, data science, and production-grade AI system engineering. Passionate about building intelligent, scalable, and high-performance AI products.

Key Skills & Technologies

- **Languages:** Python, C++, C, JavaScript, SQL

- **Frameworks & Libraries:** React Native, React.js, Vue.js, NumPy, Pandas, Matplotlib, Seaborn

- **Software Engineering:** OOP, Algorithms, Data Structures

- **Machine Learning:** Supervised & Unsupervised Learning, Time Series Analysis

- **Reinforcement Learning:** TD3, PPO, Multi-Agent RL, Reward Engineering

- **NLP:** Text Classification, Tokenization, Conversational AI

- **Cloud & DevOps:** AWS, Azure, Docker, Kubernetes, Git

- **Databases:** PostgreSQL (pgvector), Supabase, SQL Server, MongoDB

- **AI Tools:** LangChain, LangGraph, MCP, Triton Inference Server, SigNoz

Work Experience

AI Engineer, PvFritz UG
Part-Time · Remote
03/2025 – Present
Germany

- Worked on a **multi-tenant Django web application**, improving scalability, performance, and user experience through schema-aware request handling and RBAC-based access control.
- Designed and implemented a **hierarchical multi-agent AI system** (Supervisor + specialized agents) using LangChain, LangGraph, and MCP to enable intelligent, context-aware automation workflows.
- Developed a **hybrid RAG pipeline** with pgvector semantic search and Qwen3 cross-encoder reranking (85% scoring weight), enabling high-quality document retrieval across tenants.
- Integrated **Triton Inference Server** for batched embeddings and reranking, supporting multi-tenant workloads with efficient GPU utilization.
- Architected a **multi-GPU inference server** (2× RTX 5070) using Triton + OpenResty/Nginx with dynamic batching, GPU pinning, secure API keys, and rate limiting.
- Achieved **15–17 req/s throughput**, **40% latency reduction**, and **0% error rate** across 27,000+ load-test requests through systematic performance benchmarking.
- Built **15+ custom tools** for document operations, workflow execution, CRUD actions, notifications, and user/app management, integrated with tenant-aware RBAC logic.
- Implemented advanced **prompt engineering**, including modular prompt files, structured output schemas, JSON-based command execution, and fallback parsing strategies.
- Developed **real-time observability** features: message logs, agent reasoning traces, GPU metrics, error reporting, token usage tracking, and cost estimation dashboards.

Data Scientist, TAI – Turkish Aerospace Industries, Inc.   08/2023 – 10/2024
CGF Project – Autonomous Aircraft Development   *Ankara, Turkey*
- Worked on reinforcement learning and anomaly detection tasks for autonomous aircraft systems.
- Applied Isolation Forest, Local Outlier Factor, and Autoencoders on time-series anomaly detection.
- Developed RL agents for F-16, SU-27, and HURJET aircraft simulations.
- Implemented DDPG and PPO algorithms from scratch for performance benchmarking.
- Built multi-agent dogfight training environments.

Data Scientist (Part-Time), TAI – Turkish Aerospace Industries, Inc.   12/2022 – 08/2023
Working Student   *Ankara, Turkey*
- Performed anomaly detection and flight data analysis using ML methods.
- Designed reward functions to improve RL training stability and convergence.
- Built custom TD3 reinforcement learning model for control optimization.

Projects

TheUnichat – Mobile App Startup   01/2025 – 09/2025
- Founded and developed **TheUnichat**, a **React Native** mobile app for student communities.
- Built full-stack architecture using **Supabase**.
- Launched on iOS and Android, reaching **500+ users** in the first week.
- Led design, development, deployment, and marketing processes.

Movie Recommendation System   10/2024 – 03/2025
- Developed hybrid recommendation engine using **SVD collaborative filtering** and **content-based filtering**.
- Integrated real-time feedback system with **100 ms latency**.
- Ensured scalability, security, and GDPR compliance using Docker and Kubernetes.

Certificates

IELTS   09/2024 – 09/2026
Listening 7.5 · Reading 7.0 · Writing 6.0 · Speaking 7.5 · Overall 7.0 (C1)   *Ankara, Turkey*

Languages

English: C2

German: A2

Turkish: Native