Mustafa Kerem Kurban

Geneva, 1209, Switzerland• (+41) 77-927-9606
Email • Website • LinkedIn • Google Scholar

Machine Learning Engineer | MLOps & LLM Specialist

Innovative Machine Learning Engineer with 8+ years of experience leveraging cutting-edge AI tools to develop scalable systems, deploy advanced ML pipelines, and solve complex challenges in knowledge extraction, semantic search, and large-scale modeling. Proven ability to design AI-driven solutions across multiple domains. Passionate about leveraging generative AI to solve real-world problems and contribute to scaling innovative startups.

Work Experience

Neptune AI
Machine Learning Blogger 03/2025 - Current

- Contributing to the blog portal of neptune.ai with novel techniques in LLM Research, and demonstration of multiple transformer techniques (LoRA, QLoRA, DORA, MoE, SFT) on Foundation Models.

EPFL Blue Brain Project, Geneva, Switzerland
Machine Learning Engineer 04/2024-12/2024

- Developed multi-agent LLM systems using LangChain and OpenAI Swarm for query-based scientific automation, **reducing manual analysis time by 90%**.
- Designed and deployed retrieval-augmented generation (RAG) pipelines, leveraging Neo4j for semantic search and personalized knowledge graph-based recommendations, **reducing hallucinations by >90%**.
- Designed semantic analysis pipelines to deliver **personalized insights for CRM and trend forecasting**, demonstrating cross-domain adaptability.
- Optimized AI workflows with AWS SQS and vector databases, ensuring real-time updates and scalability for production environments.
- Built user-friendly, React-based interfaces for in-house testing, improving usability of advanced AI models.

EPFL Blue Brain Project, Geneva, Switzerland
Scientific Software Developer 04/2021-04/2024

- Architected high-throughput brain analysis and simulation pipelines, **processing over 100TB of multimodal neuroscientific data** to reconstruct and validate detailed neural circuits across hippocampus, somatosensory cortex, and thalamus regions.
- Led comprehensive mapping of rat and mouse hippocampal circuits by integrating neuronal morphologies, synaptic connectivity patterns, electrophysiological recordings, and molecular profiles, resulting in multiple publications.
- Innovated new approaches for analyzing multi-scale connectivity data and single cell modeling, developing unified computational frameworks to bridge micro-to-meso scale neural architectures through integrated analysis of multi-omics datasets.

Education

BSc in Life Sciences focusing on Neuroimaging and Deep Learning (2018), Bogazici University, Turkiye
MSc in Computer Science and Neuroscience (2021), Bilkent University, Turkiye / EPFL, Switzerland

Certificates

- Neo4j Certified Professional (2024)
- LLMOps and Fine Tuning Large Language Models (DeepLearning.AI, 2024/25)
- AI Agents in LangGraph (DeepLearning.AI, 2024)
- LLMs as Operating Systems: Agent Memory (DeepLearning.AI, 2025)
- Federated Fine tuning of LLMs with Private Data (DeepLearning.AI, 2025)
- AWS Foundations, AWS Bedrock (AWS, 2024)
- Infrastructure as Code in Google Cloud Platform (Linkedin Learning, 2024)
- Quantization Fundamentals with Hugging Face (DeepLearning.AI, 2025)

Technical Skills

- **Programming Languages**: Python, SQL, Bash, JavaScript, C++, Cuda, Go
- **AI and Data Tools:** PyTorch, TensorFlow, LangChain, Scikit-learn, Hugging Face Transformers, Unsloth, Peft
- **Databases:** Neo4j, PostgreSQL, ElasticSearch, Opensearch, Redis
- **Cloud Solutions**: AWS S3, EC2, SageMaker, Lambda, CloudWatch, Neo4j Aura DB/DS
- **DevOps Tools**: Docker, Kubernetes, Git, FastAPI, Terraform
- **Graph Analytics:** NetworkX, GraphSAGE, Cypher Query Language, GraphQL
- **Front-End Development:** React.js, HTML/CSS, REST APIs

Projects

Knowledge Graph Powered Chat Agent with Neo4j and LangChain:
Designed a semantic analysis pipeline integrating Neo4j and RAG to process and personalize CRM workflows for **50k+ customer records,** reducing response time by **60%.** Leveraged AWS for scalability and ensured compliance with data privacy standards. (see https://github.com/BlueBrain/citation-graph)

Agentic Workflows for Simulation Automation:
Developed a human-in-the-loop AI system using LangChain and OpenAI APIs to automate NEURON simulation workflows. Enhanced efficiency by **70%**, achieving seamless integration with existing HPC infrastructure. (https://github.com/BlueBrain/neuroagent)

Front and Backend Development for Custom Chatbot Interfaces:

Developed a personalized website using React, Node.js, and CSS, featuring blog sections with keyword-based filtering to enhance UX, and **designed a custom Chat UI** to optimize usability for in-house chatbot applications tailored to Graph RAG application.

Optimizing Large Language Models for Scientific Applications:

Enhanced OpenAI's evaluation framework for processing multiple tool calls from a single user input across complex scenarios. Addressed the low coverage rate of the baseline LLM, achieving a 20% improvement through supervised fine-tuning (SFT) on open-source and industry-standard models.

For open sourced projects, publications and additional details, visit keremkurban.created.app.