

Berke Deniz Bozyigit

Computer Scientist

berkebzt@gmail.com | +44 7552 837 419 | github.com/berkebzt | linkedin.com/in/berkebozyigit | London, UK

SUMMARY

Experienced Data Scientist specializing in large language model fine-tuning and generative AI pipeline optimization. Achieved 20% processing time reduction through parameter-efficient fine-tuning of 7B Llama-2 architecture, and optimized DreamBooth/Stable Diffusion inference pipelines to 2.1s latency via parallel processing and memory optimization. Currently researching novel nonlinear aggregation functions in neural architectures (F-Mean, Gaussian, Hybrid) using PyTorch, demonstrating improved robustness on CIFAR-10 benchmarks. Expertise in end-to-end ML deployment, OCR system optimization with Tesseract, ensemble methods (XGBoost, LightGBM), and production-scale data pipeline architecture using Python, SQL, and distributed computing frameworks.

PROFESSIONAL EXPERIENCE

Data Science Specialist, Akbank	10/2022 – 04/2024 Istanbul, Turkey
<ul style="list-style-type: none">Fine-tuned a 7B-parameter Llama-2 LLM in Python to automate HR processes, achieving a 20% reduction in candidate processing time.Implemented and evaluated Tesseract-based OCR system across multiple datasets, comparing performance metrics to optimize IBAN extraction accuracyIntegrated DreamBooth and Stable Diffusion into Streamlit pipelines, optimizing latency to 2.1s using parallel processing.Worked with Git version control and communicated deliverables across product and data teams.	
Data Science Intern, HAVELSAN EHSIM	07/2021 – 11/2021 Ankara, Turkey
<ul style="list-style-type: none">Developed and optimized Python-based data pipelines for radio frequency signal analysis, enhancing large-scale data processing throughput.Supported model optimization using Scikit-learn, SciPy, and Pandas; visualized results using Matplotlib and Seaborn.	
Data Engineering Intern, Xinerji Software	04/2021 – 07/2021 Istanbul, Turkey
<ul style="list-style-type: none">Created ETL pipelines and conducted data cleansing using SQL/PostgreSQL, supporting downstream ML model accuracy.	

EDUCATION

M.Sc., Artificial Intelligence, University Of York	09/2024 – present York, United Kingdom
<ul style="list-style-type: none">Dissertation: A New Type of Artificial Neuron: Exploring Nonlinear Aggregation Functions – designed and implemented novel neuron architectures (F-Mean, Gaussian, Hybrid) in PyTorch, achieving improved robustness on CIFAR-10 benchmarks.	
Teaching Assistant, Data Science Course, Sabanci University	07/2022 – 09/2022 Istanbul, Turkey
<ul style="list-style-type: none">Supported course delivery and prepared Python-based hands-on ML exercises.Reported technical findings to both technical and non-technical audiencesConducted literature reviews and contributed to research on machine learning methodologies.	
B.Sc., Computer Science and Engineering, Sabanci University	09/2017 – 08/2022 Istanbul, Turkey
<ul style="list-style-type: none">Graduation Project: DataSpace<ul style="list-style-type: none">DataSpace is a spatio-temporal data analytics platform project. Managed online and offline data capturing of Time and Location dimensions. Leveraged Python, MongoDB, and React for development, addressing data integration challenges across diverse sources, including climate measurements, population data, online activities, sentiment analysis.	

SKILLS

Programming: Python, SQL, Git, Linux | Frameworks: PyTorch, TensorFlow, Scikit-learn, Streamlit ML/AI: LLM fine-tuning, Generative AI, OCR, XGBoost, LightGBM, Computer Vision, Data Pipelines Languages: Turkish (Native), English (Professional working proficiency)