Farbod Farhangfar

Email: farbod.farhangfar.dev@gmail.com   Phone: +989941653285
LinkedIn: https://www.linkedin.com/in/farbod-farhangfar/   Github: https://github.com/FarbodFarhangfar

About

Data scientist and data engineer with 8+ years of experience building AI solutions across machine learning, deep learning (Generative AI, NLP, recommender systems, computer vision), ETL pipelines, data warehouses, and visualization tools. Python developer with a full-stack background (Django, FastAPI, Flutter, React) and experience building REST APIs. Hands-on with AWS, GCP, Azure, and containerization using Docker and Kubernetes. Experienced in MLOps, CI/CD, and automation with GitHub Actions. Strong collaborator in Agile Scrum teams. fluent in English, familiar with German and Turkish languages. Passionate about using data to drive real-world impact and work on large scale projects.

Experience

Senior AI product Owner   **January 2024 - March 2025**

Jarvic.ai | Dublin, Ireland

- Developed SaaS conversational agents to handle user requests and perform tasks including planning, selling, and FAQs based on 12 different b2b enterprise client's needs. Worked on its CI/CD and continues and automated operation.
- Created a generative AI engine to dynamically switch between cloud and on-premises LLMs (including open source llms), with load balancing based on business needs for chatbots and LLM apps. reduced latency by 50%, and enhanced user experience.
- Programmed RAG systems and chatbot backend with streaming, history management, and interfaces.

AI Team Lead   **February 2022 - January 2024**

Nafun Electronics (NAFUN INTERNATIONAL. LLC) | Muscat, Oman

- Managed a 5-member team and Deployed an Industrial ML Predictive Maintenance system, securing 3 contracts within the steel and oil sectors in just 6 months and Reducing maintenance costs by 5%. This system garnered substantial interest within the steel, petrochemical, and other industrial sectors that became 40% of the company's Revenue.
- Designed and implemented a high-performance Python ETL pipeline using Airflow and Spark, optimizing CPU and I/O utilization and streamlining data cleansing logic, resulting in a 30% reduction in processing time.
- Engineered a high-performance data pipeline for a SaaS platform handling 100k+ daily events using Apache Kafka and Spark, while spearheading the migration from legacy Hadoop infrastructure to a new real-time architecture.
- Collaborated in an R&D to utilize an edge-AI board to be able to run generative AI on the hardware and embedded systems.
- Constructed customizable RAG system with local and cloud LLM with approximately 10k user across 4 client businesses.

Senior data engineer / Computer vision   **March 2021 - February 2022**

Magifits | Toronto, Canada

- Evaluated computer vision algorithms in a team project, Producing new methods for generative AI virtual try-on purposes.
- Implanted backend for computer vision project with AWS. Moving the app to AWS Provided faster launch process.
- need for 3 functionality resulted in 3 different lambda functions with docker images.
- Built AWS Snowflake with automatic clustering and micro-partitioning for 50% faster image queries, while implementing real-time monitoring with CloudWatch to boost pipeline stability and troubleshooting efficiency by 40%.
- Utilized Docker and Kubernetes to handle at least 4 containers, because of different dependencies and usages.

AI engineer / Backend Developer   **December 2020 – March 2021**

Nafun Tech (NAFUN INTERNATIONAL. LLC) | Muscat, Oman

- Architected a customizable recommender system with Django Rest API for 7 different products and websites.
- Usage of Flexible Rest API reduced request size and prepared opportunities for new investors.
- Collaborated on a Chatbot project backend, to handle 5k user requests every day
- Deployed an LLM-based solution on GCP, reducing manual processing time by 60%.

Data engineer / Backend Developer   **January 2019 – December 2020**

Digikala| Tehran, Iran

- Executed scalable ETL pipelines to process millions of transactions, improving data processing efficiency by 30%.
- Maintained data pipelines using Spark and Airflow, handling high-volume structured and unstructured data.
- Built high-performance APIs and microservices using Python (FastAPI, Django) integrated with PostgreSQL, Redis, Kafka, and cloud architecture (AWS EC2, lambda, S3).

Freelancer   **September 2018 – January 2019**

self-employed | self-employed

- Research Deep Image Classification: Classify cancer images in PyTorch and OpenCV for cancer detection.
- Midi music player service: creating and playing midi music with Python.
- Spline-Visualizing-Service: display spline lines on an image using OpenCV, and Matplotlib in Flask framework.

Education

Shahrekord University - Shahrekord, Iran   **September 2017 - August 2021**

- Bachelor's degree, Computer Engineering

Skills

- Languages: Python, R Programming, C/C++, Dart, Bash
- Tools / Libraries: Scipy, Numpy, Pandas, Matplotlib, Scikit-learn, PyTorch, Tensorflow, Keras, OpenCV, RabbitMQ, ElasticSearch, haystack, LangChain, Apache Airflow, Kafka, PowerBI, kibana, Grafana, Tableau, Power BI, kibana, Git
- Frameworks: FastAPI, Django, DRF, Flask, Flutter, React
- Data Storage: MySQL, PostgreSQL, MongoDB, Redis, Vector database, AWS (RedShift, S3, DynamoDB), Azure Storage, Snowflake, Hadoop, Apache Spark
- Cloud/DevOps: Docker, Kubernetes, GCP, AWS (Glue, EC2, RDS, Lambda, Kinesis), Azure, mlflow, CI/CD, Github Actions, Linux
- Others: ETL, Web Scraping, LLM (OpenAI, Llama), Agile Development, NLP, Vector Search, RAG

Projects

AI Engine / AI Agents

- Developed an AI engine to serve multiple LLMs and embedding models for diverse use case both (SaaS and B2B), including Chatbots and RAG.
- Integrated health diagnostics and failover mechanisms to ensure seamless model switching, Improved testing, diagnostics, and reducing latency by 50%.
- Designed a load balancer to distribute queries efficiently and maintain high availability reducing latency by 40%.
- Engineered AI agents to perform tasks like event planning, calendar, and etc… based on user queries and business needs.

History Management / User usage limiter

- Implemented history management to retain user chats and key ideas, enabling the AI to provide more accurate responses.
- Built a system to manage user access based on session and request limits, enabling expirable free plans.

State manager / Call Handler

- created a state management system for a chatbot platform to ensure seamless logging and reliable user session tracking.
- Programmed an Asterisk-based phone call handling system that processes incoming calls and delivers chatbot responses.

Certificates

Deep Learning Specialization

- Improving Deep Neural Networks: Hyperparameter Tuning, Regularization and Optimization
- Structuring Machine Learning Projects
- Convolutional Neural Networks
- Sequence Models
- Neural Networks and Deep Learning

IBM Data Science Professional Certificate

- What is Data Science?
- Databases and SQL for Data Science with Python
- Tools for Data Science
- Data Analysis with Python
- Data Science Methodology
- Data Visualization with Python
- Python for Data Science, AI & Development
- Machine Learning with Python
- Python Project for Data Science
- Applied Data Science Capstone

DeepLearning.AI TensorFlow Developer Professional Certificate

- Introduction to TensorFlow for Artificial Intelligence, Machine Learning, and Deep Learning
- Convolutional Neural Networks in TensorFlow
- Natural Language Processing in TensorFlow
- Sequences, Time Series and Prediction

IBM Data Engineering Professional Certificate

- Hands-on Introduction to Linux Commands and Shell Scripting