SEP AMINIAN

Portfolio: Sephml
Linkedin: amirsepehr-aminian/ Location: London, UK
Contacts: +447878807771 amirsepehr.aminian@gmail.com

SUMMARY

Engineer building real-time, scalable ML platforms with deep expertise in high-performance data pipelines, cloud infrastructure (AWS, Azure), and model deployment. Saved £2.8m annually by developing scalable NLP solutions. Skilled in end-to-end Machine learning development, NLP, and cloud-based applications. Passionate about operational excellence in ML systems at scale, and excited by the intersection of deep learning and finance.

EXPERIENCE

Software Consultant – Assistant Manager – S&W Group Jun. 2024 – Present
London, United Kingdom

- Led requirements engineering and scoping for proof-of-concept (POC) initiatives, delivering clear timelines and execution strategies.
- Built a cloud-based ML pipeline using GPT-4 and LangGraph to automate reporting and reviews, cutting over 20 hours per engagement and saving £2.8M annually.
- Integrated Azure SQL for structured data transformation, enhancing business intelligence workflows across multiple client use cases.
- Led system design of end-to-end pipelines with traceable logs and reproducible outputs for enterprise-level audit compliance and stakeholder reviews within internal tools.
- Building concurrent data ingestion and processing pipelines handling multiple streams using multiprocessing and async Python, achieving scalable ingestion of technical documentation.
- Established automated model monitoring and performance feedback loops while advising on best practices using Python, JavaScript, and cloud-native architectures.
- Designed and deployed fault-tolerant ML systems in production with 99.9% uptime, ensuring stable performance under varying load

Software consultant - RSM UK Mar. 2023 - May 2024
London, United Kingdom

- Facilitated the incorporation of Microsoft Copilot into the business as a part of a wider team by designing detailed use case scenarios and integrating considerations on privacy policies, data sharing, as well as accuracy. This improved the overall productivity of a team of 50, increasing it by 18% based on project completion time.
- Automating manual tasks through developing robust Python scripts to increase efficiency and productivity.
- Client-facing for advisory work by holding discussions with senior technical stakeholders and drafting relevant technical documentation about projects.

Machine Learning Engineer - Definely Apr. 2022 - Mar. 2023
London, United Kingdom

- Developed NLP classifiers at scale with fine-tuned transformer models (BERT, XGBoost, and Spacy), achieving over 0.9 F1 Score on legal contract analysis.
- Led a team of 2 to engineer and implement a comprehensive text extraction pipeline using Python and Azure SQL for pre-training data preparation.
- Engineered a robust data preprocessing pipeline using Azure SQL, optimising query performance and storage efficiency.
- Spearheaded the creation of an efficient pipeline for labelling data, improving and optimising engineer productivity through labelled data management, repository oversight, and task automation.

- Engineered and monitored real-time inference pipelines with strict latency requirements, reducing average response time from 800ms to 250ms via ONNX + quantisation

EDUCATION

MSc. Artificial Intelligence with Distinction- Queen Mary University of London Sept. 2021 - Sept. 2022

- **Details**: Courses on Statistical machine learning, Reinforcement learning, NLP and computer vision.

BSc. Geomatics Engineering with Distinction- University of Tehran Sept. 2016 - Feb. 2021

- **Details**: Minor in Computer Engineering.

SKILLS

- **Programming Languages:** Python, JavaScript, C++
- **Machine Learning:** PyTorch, TensorFlow, Keras, GPU programming, LangChain, LLMs, Transformers, Scikit Learn, NumPy, Pandas, Matplotlib, NLTK, Spacy, Weights and Biases, TensorRT, TensorBoard
- **Frameworks and technologies:** Node.js, Express.js, React.js, Django, Azure cloud, AzureSQL, Redis, MongoDB, Git, Docker, CI/CD with GitHub Actions, MLFlow, Distributed Systems, Asynchronous Programming, Concurrent Execution
- **Certifications:** Intro to Software Engineering(IBM), Machine learning Summer School 2023(OxML), Neural Networks and Deep Learning(Coursera), Python for Data Science(IBM)

SELECTED PROJECTS

Planning AI Assistant Apr. 2024 – Present

- Led a team to design and deploy a scalable ML pipeline processing 1M planning application PDFs, with optimised document retrieval using Pinecone and distributed processing.
- Built a reliable data ingestion pipeline to collect and warehouse planning applications in PDF format, enabling structured extraction and downstream analysis.
- Developed a microservice using Flask, MongoDB Atlas, LangChain, and GPT-4 to extract 140+ features from PDFs via advanced prompt engineering techniques.
- Created an AI assistant capable of accurately answering planning-related queries with source references, and implemented a fast facet-based search engine using Express.js and MongoDB.

Machine Learning & Data Science Projects Aug. 2024 – Present

- Built a pipeline for recovering LLM prompts using Gemma-7B, achieving a performance score of 0.68. Applied advanced prompt engineering and evaluation techniques.
- Developed and fine-tuned neural networks for a range of NLP tasks, including language modelling, machine translation, named entity recognition, and co-reference resolution, using TensorFlow Keras and PyTorch with ELMO/BERT embeddings and attention mechanisms.
- Recreated key computer vision architectures (GoogLeNet, VGG16, ResNet) from scratch for MNIST and CIFAR-10 datasets. Also implemented SRCNN for super-resolution and GANs for image generation tasks.
- Built a real-time object detection pipeline using a fine-tuned YOLO model integrated with a live camera feed, enabling accurate classification and localisation under real-world conditions.
- Engineered a legal matter email categorisation system incorporating clustering, summarisation, classification, and tagging. Used vector-based semantic similarity and LLMs to structure loosely related threads into coherent legal cases.
- Implemented an end-to-end file type classification system leveraging pretrained models to automatically identify and label document types (e.g., invoices, bank statements, identification documents) with high accuracy across mixed-format datasets.