KALEB SOFER

kalebsofer.com
kalebsofer@gmail.com
+44 7804 333104
github.com/kalebsofer
linkedin.com/in/kalebsofer/

Machine Learning Engineer

Senior engineer with five years experience delivering scalable, production data systems. I have designed and owned upstream processes including infra and ETL, as well as user-facing ML models and data visualisation. I am adept at managing and collaborating with diverse teams, honed through consulting roles at both start-ups and established businesses.

Having recently dedicated ten weeks to training and deploying generative AI with the Machine Learning Institute, I am looking for a new environment to make an impact.

Outside of work, I am a saxophonist who enjoys spending weekends recording and producing in my studio/shed. I am also a licensed sailboat skipper, often exploring the British coast whenever time allows.

Stack

Data Engineering
Python SQL NoSQL
Spark Kafka Data quality testing (Great Expectations, YData-Profiling)

Orchestration:
Airflow Prefect

Cloud and Infra
AWS Azure GCP Terraform

ML Ops
CI/CD Docker Kubernetes

Management
Jira Confluence Trello Miro

ML Frameworks
PyTorch Scikit-learn

ML Concepts
Data preparation
Feature engineering
Regression and
Classification for
un/supervised learning

AI Architecture
CNNs RNNs LSTM GRU
Transformers

Visualisation
Plotly Matplotlib
PowerBI Excalidraw
WandB

Education

Advance AI Bootcamp, *Sep-Dec 2024*
Building, training and deploying generative AI models
The Machine Learning Institute

B.Eng. Aerospace Engineering
First Class Honours
The University of Manchester

Experience

MADETECH / Mar 2022– Apr 2024
Delivering data and digital across the public sector, I was promoted into a leadership role quickly and delivered end-to-end projects to meet the demands of evolving gov policy.

Ministry of Justice, UK / Jul 2023– Apr 2024
Senior Data Engineer
Led our data engineering team in the delivery of a **massive data migration** pertaining to court cases, subsequently **implementing a data lake**, performed **live data linkage at scale** and supported a front-end team to serve data.

AWS Deployment with Docker/Kubernetes, data lake with Apache Iceberg on AWS, PySpark pipelines, Airflow orchestration.

DVLA, UK / Mar 2022– Jun 2023
Senior Engineer
Developed an **analytics platform with Databricks**, I was responsible for **provisioning infrastructure on Azure**, integrating numerous data sources, and demonstrating the capability of our platform to build reliable and clear dashboards quickly.

Databricks on Azure, Infra with Terraform, PySpark pipelines, PowerBI dashboards.

KUBRICK / Mar 2019– Nov 2021
A training provider, delivering an intensive peer-led curriculum in preparation for consultancy work.

SANTANDER / Aug 2020– Nov 2021
Data Scientist - Financial Crime
Developed an anti-money laundering tool involving **ML methods for anomaly detection and NLP to accelerate investigations**. Deployed in production, it was used daily by **300 internal analysts.**

Backend with PySpark, modelling with PyTorch, NLP with Spacy and dashboards with Plotly.

DELIVEROO / Jun 2019– Jun 2020
Data Engineer - Payments
Business automation in response to COVID-19, including the development of a **flexible payouts platform** for restaurant payments to help with cashflow during lockdowns.

Business automation in Python/SQL, deployed with Docker, pipelines on AWS and Snowflake, Snowpipe for CI.

Check out my ongoing personal projects at kalebsofer.com