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

KALEB SOFER

Machine Learning Engineer

A senior engineer with a track record in delivering scalable and effective solutions to complex data challenges. Adept at communicating and collaborating with diverse teams, honed through consulting roles with both start-ups and established businesses. I recently transitioned into the public sector following the pandemic, bringing lessons learned from scaling businesses to improving data reliability across government.

In addition to professional work, I am passionate about music - I play saxophone and have been learning to record and produce, I also hold a sailboat skipper licence for windier weather.

Stack

Data Engineering

Machine Learning

Python SQL

PyTorch Scikit-learn

NoSQL: MongoDB

Visualization

Spark Kafka Docker Kubernetes

Plotly Matplotlib PowerBI Excalidraw

Orchestration:

Management

Airflow Prefect

Jira Confluer

Cloud and Infra

Confluence Trello

AWS Azure GCP

Terraform

Miro

Education

B.Eng. Aerospace Engineering First Class Honours

The University of Manchester

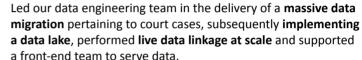
Experience

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



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

DVLA, UK / Mar 2022- Jun 2023

Senior Engineer

Development of 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 GROUP / Mar 2019- Nov 2021

A training provider and consultancy who run an intensive peer-led curriculum which I completed before working with clients.

SANTANDER / Aug 2020- Nov 2021

Data Scientist - Financial Crime



Developed an anti-money laundering tool involving machine learning for anomaly detection and NLP for press article relevancy. Deployed in production, it was used daily by 300 internal analysts.

Backend with PySpark, modelling with Scikit - learn, 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.