

# 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-ready 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 with both start-ups and established businesses. Having recently transitioned to the public sector, I have been able to bring lessons learned from scaling businesses to help improve 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 warmer weather.

## Stack

### Data Engineering

Python / SQL /  
NoSQL (MongoDB) /  
Spark / Kafka / Docker /  
Kubernetes

### Orchestration:

Airflow / Prefect

### Cloud and Infra

AWS (strong understanding)  
Azure / GCP (good  
understanding)

### ML Ops

CI/CD practices /  
Terraform /  
Data quality testing  
frameworks (Great Expectations,  
YData-Profiling)

### ML Frameworks

PyTorch / Scikit-learn

### ML Concepts

Data preprocessing  
techniques / feature  
engineering / regression  
and classification for  
un/supervised learning

Deep learning: Neural  
networks, CNN, RNN,  
LSTM, GRU, Transformers,  
image detection (YOLO)

### Visualisation

Plotly / Matplotlib /  
PowerBI / Excalidraw

### Management

Jira / Confluence /  
Trello / 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

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

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.*

Check out my ongoing personal projects at [kalebsofer.com](https://kalebsofer.com)