

# Behrad Zabihi

+44-7493-411-334 | behradz.work@gmail.com | github.com/8ehrad | linkedin.com/in/behrad-zabihi | portfolio

## Experience

- Data Scientist** Nov. 2023 – Present  
Informed Solutions Manchester, UK
  - Automated Scottish farmers' funding validation with a multimodal ML pipeline (Llama, LLaVA, Bedrock, Textract), cutting review time from 2 weeks to real-time and achieving an F1 score of 0.91.
  - Built a scalable ETL pipeline using Airbyte and DBT to extract and link EPC and postcode data for 65,000 households in the Isle of Wight, powering an interactive map for household-level EPC insights.
  - Built a Power BI dashboard enabling real-time data-driven decisions for delivery managers and recruiters.
  - Developed cloud-native web apps and REST APIs to support scalable, data-driven public services.
- Data Science Placement/Dissertation** Jun. 2023 – Sep. 2023  
Informed Solutions Manchester, UK
  - Developed an NLP pipeline for keyword extraction and summarisation of structured and unstructured data.
  - Improved topic extraction using BERTopic and LLM-based techniques.
  - Used GPT and LangChain to summarise urban planning cases, cutting manual review effort and reaching a BERTScore F1 of 0.84.
- Data Scientist** Jun. 2022 – Aug. 2022  
Mizan Gostaran Sharif Tehran, Iran
  - Scaled resume parsing system from 26K to 500K+ CVs using a custom NLP pipeline for entity extraction.
  - Improved job matching by clustering unstructured CVs using UMAP + HDBSCAN for semantic similarity analysis.
- ML Research Assistant** Feb. 2021 – Jul. 2021  
Sharif University of Technology Tehran, Iran
  - Modelled Covid-19 transmission patterns across countries using time series and statistical modelling on 100k global cases.
  - Identified key influencers in the global propagation graph by inferring inter-country transmission rates.

## Selected Projects

- Bias in Fake News Detection** | *NLP, LSTM, Python, PyTorch* Mar. 2023 – May 2023
  - Investigated LSTM model vulnerability to spurious correlations in fake news classification by injecting synthetic bias into training data.
- Movie Sentiment Analysis Pipeline** — *Apache NiFi, Python, SQL, Grafana* Mar. 2023 – Apr. 2023
  - Designed a near-real-time, containerised data pipeline using Apache NiFi and Python (nltk) to ingest and analyse IMDb movie review sentiment.
  - Visualised sentiment trends in Grafana to explore correlations with ratings and box office performance.
- Co-Op Bakery Daily Forecasting** | *Time-Series Forecasting, Optimisation, Python, R* Dec. 2022 – Jan. 2023
  - Trained time-series models (GLM, LSTM, SARIMA) to forecast daily bakery demand, achieving an  $R^2$  of 0.67.

## Education

- Lancaster University** Lancaster, UK  
MSc in Data Science, **Distinction (top 2.5%)** Oct. 2022 – Sep. 2023
- Sharif University of Technology** Tehran, Iran  
BSc in Computer Engineering Sep. 2016 – Jul. 2021

## Skills

**Programming Languages:** Python, C/C++, R, SQL, C#  
**Machine Learning:** TensorFlow, Keras, PyTorch, Scikit-learn, Pandas, NumPy, SciPy, Huggingface  
**Visualisation:** Power BI, Matplotlib, Seaborn  
**Technologies/Frameworks:** AWS, Azure, Git, Docker, Hadoop  
**Languages:** English, French, Persian