Kostas Stathoulopoulos

k.stathou@gmail.com
+44(0)7783530330
https://kstathou.github.io/

Work Experience

Lead Machine Learning Engineer 11/2020 - now
Util

I lead ML at Util. I designed and built Util's ML-driven knowledge graph. I developed end-to-end ML pipelines to extract products and sustainability concepts from text data and measure the social and environmental impact of publicly traded companies.

Achievements:
- Developed and deployed deep learning models (transformers) for entity recognition, entity linking and relation extraction.
- Built Util's MLOps infrastructure for the ML lifecycle and model deployment using Weights & Biases, Docker and GitHub Actions.
- Orchestrated all ML jobs in Airflow and used Spark (Databricks) for data processing and batch inference pipelines.
- Built a bottoms-up product ontology by using Large Language Models (LLMs) to extract products from company websites, financial reports and Wikidata and organise them hierarchically.
- Designed Util's data model for NLP data and implemented it in dbt.
- Collaborated with the Product team on the ML Roadmap and managed the ML team.

Generative AI Resident 07/2023 - 10/2023
Nesta

I investigated how Generative AI, specifically LLMs, can be used for social good. I explored the potential of LLMs in early years education and shared good AI engineering practices across the organisation.

Achievements:
- I delivered five open-source, AI-based prototypes, a series of technical blogs on LLMs, Retrieval-augmented Generation (RAG), vector databases and how to use OpenAI's models.
- I built a system that logs user conversations and feedback and evaluates the performance of the AI prototypes.
- I mentored Nesta's Data Science team on AI engineering practices through 1-1s, office hours and presentations on MLOps.

Machine Learning Engineer 10/2019 - 10/2020
Mozilla

I joined Mozilla as an Open Science Fellow and led a remote team of a data visualisation engineer and a UX researcher. We built Orion, a knowledge discovery and research measurement tool for scientific texts. Orion enables users to monitor progress in science, visually explore the scientific landscape and search for relevant publications.

Achievements:
- Developed an open source data collection, enrichment and analysis pipeline for scientific data.
- Built a search engine that combines Elasticsearch for short-text queries and a vector-based search for long-text queries (Transformers and Faiss).
- Presented my work to conferences and workshops. You can find all presentations and tutorials here.

Senior Data Scientist 02/2019 - 10/2019
Data Scientist 01/2017 - 01/2019
Nesta, London, UK

I worked at the intersection of machine learning, economics and policy. I developed tools that helped policymakers make evidence-based decisions. I focused on natural language processing tasks such as topic modelling, text classification, named entity recognition and search.

Achievements:
- Developed an information retrieval system that uses word embeddings for query expansion and ElasticSearch for full-text search.
- Combined web with survey data to predict the size of the UK's Immersive Economy for the Innovate UK KTN.
- Used users' activity on GitHub to nowcast European innovation indicators.
- Presented our work to policy makers and technical audiences at conferences, panel discussions and workshops.

Junior Data Scientist 07/2016 - 01/2017
Digital Catapult, London, UK

I did my MSc Thesis project on improving the UK's industry classification system. I scraped company websites, created company vectors with doc2vec and assigned them to market sectors with fuzzy clustering. Using the dense company vectors, I prototyped a tool to find similar companies given a company name. This assisted Digital Catapult employees in discovering UK-based companies in areas like AI, VR and IoT that were not captured by the UK's industry classification system.

Teaching Assistant - Data Science 02/2016 – 05/2016
General Assembly, London, UK

I assisted the instructor in supervising and grading students. I prepared Jupyter Notebooks that showcased how to use Python packages for data science like Pandas, scikit-learn and Numpy and guided students through their course projects.

Education

2015-2017 **MSc Data Science, City University London**
Key modules: Reinforcement learning, big data, machine learning, data visualisation, neural networks.
MSc Thesis project: Mapping without a map: Exploring the UK business landscape with text embeddings.

2008-2015 **BSc Economics and Computer Science, National and Kapodistrian University of Athens**

Skills

Databases	SQL, Snowflake, dbt
Machine learning	Full Python stack (scikit-learn, pytorch, pytorch-lightning, transformers, Langchain)
Experiment tracking	Weights & Biases
Deployment	Airflow, Docker, Databricks, Github Actions
Collaboration	git
Cloud Computing	AWS, modal

Pro bono work & training

2023 Completed CoRise's **Search Engineering with ML** course
2021 NLP expert on the steering committee of **DataKind's** DataCorps projects.
2020 Mentored Kurdish students on python and data science through the Level Up programme
2019 **Mozilla Open Leaders**: Received mentorship and training through Mozilla on designing and building open, transparent and inclusive projects that support the Internet Health movement.
2018 **Deep Learning Specialisation, Coursera**

Selected publications

Mateos-Garcia, J. and Stathoulopoulos, K., Klinger, J. 2020. A narrowing of AI research? Research Policy: Special Issue: The Governance of Artificial Intelligence (also on https://arxiv.org/abs/2009.10385)

Jurowetzki, R., Hain, D. S., Mateos-Garcia, J., Stathoulopoulos K. 2020 The Privatisation of AI Research: Causes and Consequences. https://arxiv.org/abs/2102.01648

Stathoulopoulos K. 2019 An analysis of the informal knowledge network in Sub-Saharan Africa. IC2S2.

Stathoulopoulos K., Mateos-Garcia, J. 2019. Gender Diversity in AI Research. SSRN: https://ssrn.com/abstract=3428240

Klinger, J., Mateos-Garcia, J. and Stathoulopoulos, K., 2018. Deep learning, deep change? Mapping the development of the Artificial Intelligence General Purpose Technology. arXiv preprint arXiv:1808.06355.

Stathoulopoulos, K. and Mateos-Garcia, J., 2017. Mapping without a map: Exploring the UK business landscape using unsupervised learning, Data for Policy Conference