

Kostas Stathouloupoulos

k.stathou@gmail.com

+44(0)7783530330

[linkedin.com/in/kostas-stathouloupoulos-36867b74](https://www.linkedin.com/in/kostas-stathouloupoulos-36867b74)

<https://github.com/kstathou>

Work Experience

Lead Machine Learning Engineer Util

11/2020 - now

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.

Responsibilities:

- Develop and maintain a Snowflake database using dbt.
- Schedule Airflow DAGs for data processing and batch inference on Databricks.
- Develop and deploy transformers-based models for entity recognition, entity linking and relation extraction.
- Build Util's MLOps infrastructure, from experiment tracking to deployment.
- Use LLMs (GPT-4) for text data annotation.
- Use LLMs (GPT-4) to build a product ontology by extracting entities and revenue streams of public companies from financial reports.
- Collaborate with the Product team on the ML Roadmap and lead a team of an ML Engineer, a Quantitative Analyst and an Analytics Engineer.

Machine Learning Engineer Mozilla

10/2019 - 10/2020

I joined Mozilla as an Open Science Fellow and led a remote team of a data visualisation engineer and a UX researcher. We built on [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](#).

Principal Researcher Researcher Nesta, London, UK

02/2019 - 01/2020

01/2017 - 02/2019

Worked at the intersection of machine learning, economics and policy making. I developed tools that help policy makers take evidence based decisions.

Achievements:

- Developed an information retrieval system that uses word embeddings for query expansion, TF-IDF to filter terms and Elasticsearch for full-text search.
- Combined web with survey data to predict the size of UK's Immersive Economy.
- 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 Digital Catapult, London, UK

07/2016 - 01/2017

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.

Teaching Assistant - Data Science General Assembly, London, UK

02/2016 – 05/2016

I assisted the instructor in supervising and grading students. I helped the students with coding issues and explained how various machine learning models work. I also provided them with practical tutorials using Jupyter Notebooks and guided them through their individual course project.

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	Python stack (for example, scikit-learn, pytorch, pytorch-lightning, transformers)
Experiment tracking	Weights & Biases
Deployment	Airflow, Docker, Databricks, Github Actions
Collaboration	git
Cloud Computing	AWS

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