

CILLIAN BERRAGAN

Liverpool, United Kingdom L22 8QD

cjberragan@gmail.com ♦ 07787796620 ♦ linkedin.com/in/cjberr ♦ github.com/cjberr

Machine Learning Engineer at the Consumer Data Research Centre (CDRC), with over 5 years of experience programming in Python. PhD with published research papers in geographic information science and NLP. In my current role, I worked with Llama Index and LangChain to build a Retrieval Augmented Generation (RAG) LLM-powered search system for the CDRC Data Catalogue. Extensive experience with PyTorch (Lightning), Hugging Face transformers, and the OpenAI API. MLOps and data engineering with DVC, Dagster, GitHub Actions and Docker/Podman. Machine Learning and data processing with Scikit-Learn, Pandas and Polars.

WORK EXPERIENCE

Consumer Data Research Centre

Liverpool, UK

Machine Learning Engineer

Dec 2023 – Oct 2024

- Developed a Retrieval Augmented Generation (RAG) Search System for the CDRC Data Catalogue, powered by Llama Index and OpenAI, to enhance data discovery and access.
- Generated data schemas for data products to enable automated checks on ingestion of new data products; using Pydantic, Pandera (pandas/polars dataframes) and Dagster for monitoring.
- Automated the generation of a large-scale geographically aligned dataset from the common crawl using Dagster for pipeline monitoring, and Hugging Face transformers for named entity recognition + classification.

Consumer Data Research Centre

Liverpool, UK

Data Scientist (Part Time)

Feb 2021 – May 2022

- Developed a system to calculate access to amenities using the road network across each postcode in Great Britain.
- Improved calculation times from previous implementations using GPU accelerated (cugraph) graphs, reducing time from 8 hours per calculation, to 10 minutes.

TEP - The Environment Partnership

Warrington, UK

Graduate Ecologist

May 2018 – Sep 2018

- Led projects to support planning and development across the country.
- Wrote professional reports and collaborated as part of a large team.
- Travelled to sites to engage in-person with clients and site managers.

EDUCATION

University of Liverpool

Liverpool, UK

Doctor of Philosophy in Natural Language Processing & Geography

Sep 2019 – Dec 2023

- **Thesis:** Exploring Place from the Perspective of Informal Social Media Text
- Published peer-reviewed research in reputable journals and presented at international conferences.
- Extensively used Python for deep learning through PyTorch (Lightning) and Hugging Face Transformers, ensuring reproducible analysis through Docker/Podman, DVC and Git.
- Engaged with lab activities; hosting and contributing to weekly meetings. Showcasing research and technical demonstrations.

University of Liverpool

Liverpool, UK

Master's of Science in Geographic Data Science (Distinction: 82%)

Sep 2018 – Sep 2019

- **Dissertation:** Utilising Supervised Parametric Classification to Assess the Quality of the UK Rural Road Network using Aerial LiDAR Data
- Completed modules in both Computer Science and Applied Social Science covering Machine Learning, statistics and geographical analysis.
- Highest degree grade in the cohort.

University of Hull (Scarborough Campus)

Scarborough, UK

Bachelor's of Science in Marine Biology (First: 71%)

Sep 2014 – Sep 2017

SKILLS

- **Soft Skills:** Public Speaking, Communication, Critical Thinking, Leadership, Academic Writing, Technical Writing
- **Tools:** Python, Linux (Bash, Zsh, Arch Linux), Git, Docker/Podman, Neovim, Typst/LaTeX/Quarto
- **Machine Learning & Data Science:** Natural Language Processing, Deep Learning, Data Analysis, PyTorch (Lightning), Polars, (Geo) Pandas, Hugging Face (transformers, datasets), LlamaIndex/LangChain, spaCy, SciKit Learn
- **MLOps:** SQL (duckdb, maria, postgresql), Pytest, FastAPI, Dagster, Dask, DVC, GitHub Actions