

Meg Patakota

[in LinkedIn](#) | [07873956643](#) | [Portfolio](#) | [Email](#) | [GitHub](#) | [Oxford AI Product](#)

Experience

Principal Data Scientist & Manager

[Mishcon De Reya](#)

Hybrid, London, UK

03/2022 - Current

Leadership & Strategy

- Leading data science and analytics on the business operations side, serving as point person to **C-Suite**, Partners, and Heads of Departments and presenting data-driven insights that influence executive decision-making and shape company-wide strategies.
- Engaging directly with C-Suite executives to design interactive **dashboards** and **predictive models** and **leading** the firm's analytics and reporting initiatives by engineering the **KPI** reporting system using **Power BI** and **Python**.
- Architecting** the firm's **business intelligence** initiative towards centralised data, upskilling employees to self-serve Power BI analytics.

Data Science & Analytics

- Building & **leading** an end-to-end **machine learning** pipeline, from research through to deployment, to predict the time spent on new matters using **grading boosting**, employing training best practices i.e., **cross validation**, **hyperparameter tuning** and **feature engineering**.
- Developed and automated ETL pipelines using Synapse Notebooks and Apache Spark Pools** to extract data from **APIs** (e.g., Thomson Reuters HighQ), perform data processing and cleansing tasks, and load data into the bronze layer of the **Medallion architecture**. Leveraged parameterized pipelines for dynamic data handling and optimised Spark jobs to ensure scalable and efficient data processing.
- Managing a **RAG** project on relational databases so users can chat with Power BI **semantic models** to draw quick and accurate insights.
- Spearheading** the development of a **marketing analytics** framework, to integrate and analyse client data across various platforms using **statistical models** and **ML** techniques with the objective to improve customer acquisition, retention rates.
- Strong **SQL** experience working with data from 10+ diverse sources including financial, accounting, HR, sustainability, marketing, client, and legal time recording systems across 5+ databases, 50+ tables and 15+ million data points for data science and analytics.

Data Engineering & Governance Collaboration

- Optimising **architecture** and **ETL** processes, resulting in over 10% improved pipeline efficiency in the last 1 year.
- Developing data governance mechanism, a **Python**-based algorithm in **Synapse Analytics** with **Spark** including monitoring data across 5+ domains flagging 1000+ discrepancies, increasing data **accuracy** by ~30% and cutting review time by ~20%.
- Pioneering** the development and implementation of comprehensive **data models** (eg: Finance, R&C, ESG, HubSpot, HighQ), aligning data science capabilities with corporate objectives and enhancing the firm's position in responsible business practices.

AI Researcher

[University of Oxford](#)

Remote, London, UK

09/2023 - Current

- Building a **GraphRAG** chatbot focused on enabling researchers to interrogate a corpus of fossil fuel industry documents.
- Represented the corpus as a knowledge graph in AWS **Neptune**, combined with **Pinecone** for hybrid search (semantic and keyword), exposing it to a chain-of-thought LLM agent to generate query plans and execute queries to retrieve content.
- Engineered a **GitHub Actions CI/CD** pipeline for **Docker** deployments to **AWS ECR**, for an embedding model endpoint in **Lambda**, and a **Streamlit** app on **ECS**.
- Orchestrated a data pipeline to prepare and ingest data, leveraging **AWS Textract's** OCR, **GLiNER** for entity extraction, **SpaCy** for paragraph detection, **SentenceTransformers** for embeddings, and **Scikit-Learn** for paragraph clustering.
- Elevated system resilience by integrating AWS **Cognito** for secure authentication and **DynamoDB** for durable conversation persistence.
- Collaborating with Oxford Climate Litigation Lab to conduct comprehensive system evaluation and co-author pivotal **research papers**.

Education

Skills

MSc Business Analytics University College London, UK Grade: Distinction	2020-21	Artificial Intelligence	Machine Learning (Scikit-Learn, XGBoost, LightGBM), NLP (hugging face, embeddings), RAG (LangChain, Pinecone)
		Cloud & Big Data	AWS (ECS, ECR, EC2, Cognito, DynamoDB, ACM, Lambda), Azure (Synapse Analytics, Data Factory), Big Data (Spark)
MSc Accounting & Finance University of Bristol, UK	2019-20	Business Intelligence	Pandas, NumPy, PySpark (Spark SQL) NetworkX, Plotly, Matplotlib, Power BI
BSc Business Administration University at Buffalo, USA	2015-19	Web Development & Version Control	HTML, CSS, Django, Flask, Tailwind Git Docker

Data Projects

Breathe London Data Access Tool - Python Wrapper

Simplifies access to Breathe London air quality data, enabling developers to easily integrate this valuable information into their applications without writing complex API logic from scratch

Skills: API Development & Integration, Data Automation & Workflow Optimization, Open-Source Tool Development & Maintenance

CompanyMatch: Intelligent Company Name Reconciliation

Bridging the gap between diverse company databases and official registries with advanced matching & AI.

Skills: Advanced Data Matching & Fuzzy Logic, Data Embedding Techniques, Data Quality Improvement & Consistency