Danny James **Williams**

dannyjameswilliams.co.uk
danny@dannyjameswilliams.co.uk
github.com/dannyjameswilliams
linkedin.com/in/dannyjameswilliams

I want to use my knowledge in statistics, machine learning and AI for the betterment of our society, and be at the forefront of these rapidly evolving technologies. I'm interested in all fields of AI, with experience in embedding models, information retrieval, RAG, agentic systems and the applied use of LLMs.

Education

PhD (Statistics/ML/AI)
2019-2024 | **University of Bristol**

MMath Mathematics (1st)
2015-2019 | **University of Exeter**

Skills

General
- Statistical Theory
- Machine Learning · AI · LLMs
- Fine-tuning · RAG
- Prompt Engineering
- Agentic AI · Optimisation
- Embeddings · Quantisation

Programming
- Python · FastAPI · R
- LaTeX · MATLAB · PyTorch
- NumPy · SciPy
- pandas · DSPy

Notable Projects

weaviate/elysia

An Agentic AI platform for automatic tool calls using a decision tree structure. Python package and web app. More than 30k downloads on PyPI and nearly 2k stars on GitHub. *Independently constructed the backend in Python.*

content generator

An in-house LLM powered social media post writer, designed to maximise engagement and mimic author style, via prompt engineering and finetuned models.

other

For fun, I trained an image detection model for detecting anime images, as well as fine-tuned an LLM to mimic my friend's writing style. I also embedded and visualised chapters from the Game of Thrones books.

References

Available upon request.

Relevant Employment

Machine Learning Engineer | **Weaviate (Labs)** | **Feb 2026 - Present**

- A primary developer for the Query Agent backend; Weaviate's paid agentic RAG service.
- Experimental research on new methods and models.

Machine Learning Engineer | **Weaviate (Developer Growth)** | **Aug 2024 - Feb 2026**

- Developed demo applications including a popular agentic RAG app.
- Lead internal education about the latest AI research developments.

Research Intern | **Adarga** | **Oct 2021 - Mar 2022**

- Used NLP and other methods to explore embeddings of AI generated text.
- Worked on an unsupervised clustering technique for disinformation detection for AI safety.

Tutorial Leader | **University of Bristol** | **Sept 2020 - Apr 2022**

- Led tutorials for undergraduate students in Probability and Statistics courses.
- Nominated for the Bristol Inspiring and Innovative Teaching Award.

Research

Late Chunking: Contextual Chunk Embeddings Using Long-Context Embedding Models | **Arxiv (2025)**
Embeddings · Fine-tuning · Chunking · RAG · Experimentation

High-dimensional Differential Parameter Inference in Exponential Family using Time Score Matching* | **AISTATS 2025**
Unnormalised Density Estimation · Experimentation · Asymptotics

Querying Databases with Function Calling | **Arxiv (2025)**
Agentic AI · LLMs · RAG · Vector Databases

Approximate Stein Classes for Truncated Density Estimation* | **ICML 2023**
Model Fitting · Statistical Theory · Unnormalised Density Estimation

Score Matching for Truncated Density Estimation on a Manifold* | **TAG in ML, ICML 2022 Workshop**
Manifold Density Estimation · Statistical Theory · Benchmarking

Estimating Density Models with Truncation Boundaries using Score Matching | **JMLR**
Model Fitting · Unnormalised Density Estimation · Experimentation

*=first author

Other Employment

Customer Service Assistant | **Libraries Unlimited** | **2018 - 2019**

Customer Consultant | **Dixons Carphone** | **2016 - 2017**

Customer Assistant | **Tesco** | **2014 - 2015**