Dr William Mayes

Williammayes.com · +44 7592 762562 · williammayes5@gmail.com · github.com/wpmayes

Data Scientist with a PhD in Psychology and 4+ years of applied experience at the Care Quality Commission, spanning data science, survey research, and AI evaluation. Skilled in designing and executing structured experimentation with LLMs, including bias testing, output evaluation, and statistical validation. Independently developed EvalLayer, a configurable LLM evaluation framework implementing deterministic, normalized, and LLM-as-judge validation, with confidence intervals, consistency scoring, and paired model comparison. Experienced in prompt engineering, structured output design, and translating research insights into deployable systems. Doctoral training in experimental design and Bayesian modelling underpins a hypothesis-driven, rigorous approach to AI evaluation and behavioral research. Seeking to apply this expertise to product environments and scalable AI systems.

CORE TECHNICAL SKILLS

AI Evaluation & Responsible AI: Bias evaluation across protected characteristics, counterfactual test design, baseline construction, metric selection, iterative validation, LLM-as-judge chaining, structured output design, PII redaction (Microsoft Presidio), adverse impact analysis

Statistical Methodology: Bayesian MCMC, experimental design, causal inference, psychometric analysis, sampling and weighting, regression, classification, time-series modelling

Python & ML: scikit-learn, XGBoost, neural networks, Naive Bayes, SVM, pandas, numpy, statsmodels, Hugging Face transformers, FastAPI, Azure ML, TimeGPT

Data Engineering: R (package development, reproducible analytical pipelines), SQL, Git, API integration, SPSS migration

Communication & Delivery: Senior stakeholder presentation, publication authorship, cross-government collaboration, line management, technical mentorship

PROFESSIONAL EXPERIENCE

Data Scientist / Senior Researcher   Sep 2023 – Present
Care Quality Commission

– Executed structured bias evaluation of LLM outputs for patient harm classification, designing counterfactual test cases and identifying statistically significant sex bias. Implemented PII redaction via Microsoft Presidio to ensure model-safe processing.
– Led safeguarding free-text classification project, evaluating decision trees, random forests, XGBoost, and neural networks with careful cost-weighting of false negatives.
– Built serious-injury free-text classifiers; tested Naive Bayes, SVM, logistic regression; deployed models systematically.
– Managed national NHS patient surveys, including Maternity, Adult Inpatient, and Children & Young People surveys. Designed sampling, contact strategy, and weighting methodologies and oversaw the end-to-end research process.
– Led internal AI governance project: catalogued AI tool usage, assessed risk, and drafted responsible use guidelines. Delivered seminars on responsible AI, structured trials of GitHub Copilot, and line-managed junior researchers.

Researcher   Jul 2022 – Sep 2023
Care Quality Commission

– Translated legacy SPSS pipelines into reproducible R/Python workflows; maintained internal R package and established team coding standards.
– Designed and implemented a selection weighting methodology for an ethnic minority respondent sample boost; ran simulations modelling outcomes under different assumptions; made a data-driven recommendation to senior colleagues. Analysis revealed health inequalities masked by broad ethnic category aggregation, findings that were shared with NHS England and the cross-government survey research group.

Consultant   Feb 2022 – Jun 2022
Open University

– Advised methodological design of a study examining how movement of people with DCD is perceived by those with ASD; ensured diagnostic criteria alignment with DSM standards; recruited participants through DCD community groups.

INDEPENDENT PROJECTS

EvalLayer, LLM Evaluation Framework *Feb 2026 – Present*
evallayer.netlify.app · github.com/wpmayes

– Configurable framework for structured LLM output evaluation: deterministic, normalized, and semantic (LLM-as-judge) validation. Statistical layer includes Wilson score CIs, Bernoulli consistency signal, McNemar's test for paired comparison.
– Deployed React/TypeScript frontend on Netlify; FastAPI backend on Railway; supports HuggingFace Router, OpenRouter, Ollama.

Common Ground, Multi-Agent LLM Game *2025 – Present*
Private commercial project

– Proof-of-concept negotiation game in Godot with a three-model LLM pipeline for generation, state-tracking, and content safety evaluation.
– Custom Godot-to-LLM bridge enabling offline play via quantized GGUF models.

EDUCATION

PhD, Psychology *2018–2022*
University of Surrey

Thesis: The nature and influence of sensory processing deficits in Developmental Coordination Disorder. Multimodal neuroimaging (MRS, fMRI), psychophysics, Bayesian computational modelling. First evidence of GABAergic dysregulation in DCD.

MSc, Research Methods in Psychology (Distinction) *2017–2018*
University of Surrey

BSc, Psychology (Upper Second-Class Honours) *2012–2016*
University of Surrey

PROFESSIONAL DEVELOPMENT

Data Science Campus Graduate Programme *2023–2025*
Office for National Statistics

Two-year programme covering Python, SQL, ML, NLP, Azure/AWS, RAPs, data science ethics, agile project management, and applied mathematics for ML. Selected alongside a small cohort from across government.

SELECTED PUBLICATIONS

Mayes, W. P., et al. (under review). Sensory atypicality in DCD is associated with cortical excitatory-inhibitory imbalance. Journal of Neurodevelopmental Disorders.

Mayes, W. P., Gentle, J., Ivanova, M., & Violante, I. R. (2024). Audio-visual multisensory integration and haptic perception are altered in adults with DCD. Human Movement Science. https://doi.org/10.1016/j.humov.2024.103180

Mayes, W. P., Jansari, A., & Leonard, H. C. (2023). Exploring executive functioning of adults with probable DCD. Developmental Neuropsychology. https://doi.org/10.1080/87565641.2023.2264424

Mayes, W., et al. (2021). Top-down inhibitory motor control is preserved in adults with DCD. Developmental Neuropsychology. https://doi.org/10.1080/87565641.2021.1966431