

William Gitta Lugoloobi

 Github |  wlugoloobi |  williamlugoloobi.com |  williamgitta@gmail.com

Rhodes Scholar and Oxford PhD student researching safety and efficiency trade-offs in post-trained LLMs for agentic tasks. My [recent work](#) shows that LLMs encode their own likelihood of failure before generating a response, a signal I use to route queries across models, improving performance while cutting costs by 70%.

EDUCATION

University of Oxford

DPhil Social Data Science

Oct. 2024 – Present

- **Thesis:** Rewarding Reasoning: Post-Training Choices and Security Trade-offs in Agentic LLMs.
- Supervised by Prof. Joss Wright and Prof. Chris Russell.
- Elected MCR President of Keble College, representing graduate students to college governance.

University of Oxford

MSc Social Data Science

Oct. 2023 – Aug. 2024

- Courses included Applied Machine Learning, Natural Language Processing, Data Analytics at Scale, Frontiers of Data Science, and Applied Analytical Statistics.
- Fully funded by the Rhodes Scholarship.
- **Thesis:** *From Headlines to Stories: Utilising Transformers for Story Driven OCR of Newspapers.*
- Thesis commended for operationalising traditional content analysis as a Document Layout Analysis task.

Northwestern University

BS Journalism and Strategic Communications

Aug. 2019 – May. 2023

- Graduated Summa Cum Laude.
- Founded the University's first data science club and lobbied for uptake in CS and AI courses in the journalism department.

TEACHING

University of Oxford, Oxford Internet Institute

Jan 2025 — Present

- Teaching Assistant, *Introduction to Natural Language Processing* (Oxford's core NLP course).
- Led weekly tutorials on Naive Bayes, RNNs, and Transformers with over 40 MSc/PhD students.
- Designed and delivered mini-lectures on Weights & Biases for experiment tracking, re-implementing GPT-2 locally, and using shared GPU clusters (SLURM/vLLM).
- Received commendations from the course convenor and students for clarity in lectures, plus for introducing practical tooling.

WORK EXPERIENCE

Software Engineering Intern, Qatar Computing Research Institute

Apr 2022 — Jul 2022

- Prototyped and deployed a Python/Docker image-classification model to detect natural disasters from social-media images.
- Placed 3rd out of 50 interns in QCRI's Summer Intern Research Competition.
- Guided researchers in porting existing code into a production-ready application.

Data Scientist/Analyst Intern, VICE Media Group

Feb 2022 — Apr 2022

- Built a data pipeline for VICE Arabia's digital content using Python (Beautiful Soup) and external APIs.
- Designed a dashboard for campaign performance tracking and benchmarking across teams.

Data Science Intern, Al Jazeera Media Group

Apr 2021 — Aug 2021

- Implemented a deep-learning model (TensorFlow, Azure Databricks) to detect manipulated images and fake news from social media.
- Created mockups and front-end designs for a newsroom "breaking news" dashboard to streamline information from multiple sources.
- Evaluated Al Jazeera's newswires and identified undelivered articles to the company's Microsoft SQL Server database.

PROJECTS

Trajectory Kit

[Repo](#)

Open-source agentic deep research harness achieving state-of-the-art performance on DeepResearch-Bench (55.08 at release). Novel test-time verification pipeline, modular multi-agent architecture and full trace visibility.

SOTA Verl GRPO Implementation

[PR](#)

Contributed a GRPO training script to Verl, exceeding SOTA-level MATH500 for a 1.5B model with a single A100 in under an hour.

Berserk MCP

[Repo](#)

Created an MCP Server and Client with FASTMCP and VLLM to give LLM Agents the ability to search the web for information on when a new chapter of Berserk (my favourite manga) is released

PUBLICATIONS

Lugoloobi, William, Thomas Foster, et al. (2026). "LLMs Encode Their Failures: Predicting Success from Pre-Generation Activations". In: Accepted to the ICLR 2026 Workshop on Latent and Implicit Thinking (LIT). arXiv: [2602.09924 \[cs.LG\]](#).

Lugoloobi, William and Chris Russell (Oct. 2025). "LLMs Encode How Difficult Problems Are". In: arXiv:2510.18147 [cs]. DOI: [10.48550/arXiv.2510.18147](#). URL: <http://arxiv.org/abs/2510.18147>.

SKILLS

Research & ML	PyTorch, HuggingFace, VERL, GRPO, LoRA/PEFT, RLHF, reward modeling, post-training, activation probing, behavioural fingerprinting, LLM evals, red-teaming, RAG
Engineering	Python, R, TypeScript, vLLM, SLURM, Docker, Weights & Biases, Azure Databricks, React.js, Tailwind
Leadership	Keble MCR President, Northwestern Data Science Club Founder, Rhodes Scholar