WILLIAM LEENEY

Artificial Intelligence PhD

will.leeney@gmail.com Available now London willleeney.com

EMPLOYMENT

StackOne | Jul 2025 - Present

AI Research Engineer (Oct 2025 - Present)

- **Finetune small to medium language models** for **PII redaction**, **semantic search** and **tool calling**, generating synthetic training data with branching methods using **Unsloth, vLLM and Modal**.
- **Took the initiative** to build the **infrastructure for gathering data** to enable **continuous / online learning**.
- Built **centralised benchmarks** and **evaluation methodologies** to compare our **MCP servers** against competitors and direct how we built them out.
- Built an **AI agent** to research and prototype new integration use-cases.
- Developed the features and infrastructure for **discode**, a sandboxed code-execution layer that cut LLM context usage by ~**96%** for MCP tool calls.

AI Engineer (Jul 2025 - Oct 2025)

- Researched and built an **AI agent** that translates provider errors into clear, actionable resolution steps, taking an **evals-first approach** from day one.
- Designed the evaluation pipeline, created custom tools, integrated into the product and built monitoring dashboards.

OKKO Health | Jun 2024 - Jun 2025

ML Engineer

- **Built predictive models for early disease detection** from remote patient monitoring data for AMD macular degeneration patients.
- Developed **time-series training partitions** for variable-length patient sequences and built an **anomaly detection** system, deploying a hyperparameter optimization and evaluation framework with **PyTorch**.
- Communicated complex technical concepts simply to non-technical stakeholders and ensured compliance with pertinent AI regulations.

SKILLS

• **Python** • TypeScript • Go • Swift • Julia • C • Bash • SQL • **PyTorch** • **LLM Finetuning** • Hugging Face • **Transformers** • TRL / PEFT • LoRA • **vLLM** • Unsloth • TensorFlow • **JAX** • scikit-learn • Optuna • NumPy • SciPy • Pandas • Numba • WandB • **CUDA** • Metal • MLX • Modal • HPC • AWS • Cloudflare • Vercel • Docker • Terraform • Dagster • React • Next.js • Astro • Tailwind • FastAPI • Flutter • SwiftUI • Git • MCP • Claude Code • LaTeX • HTML • CSS

PUBLICATIONS

- *Dingeto and **Leeney**, "AgentRedBench: Dynamic Redteaming and Integration-Aware Defense for LLM Agents over SaaS Integrations", arXiv, 2026.*
- ***Leeney** et al, "Non-Dissipative Graph Propagation for Non-Local Community Detection", IJCNN, 2025.*
- ***Leeney**, Rathore and Campbell, "Automated AMD Vision Event Flagging Using Machine Learning with Remote Vision Monitoring", Investigative Ophthalmology & Visual Science, 2025.*
- ***Leeney** and McConville, "Uncertainty in GNN Learning Evaluations: The Importance of a Consistent Benchmark for Community Detection.", Complex Networks, 2024. **(Invited Oral)***
- ***Leeney** and McConville, "A Framework for Exploring Federated Community Detection", AAAI GCLR, 2024.*
- ***Leeney** and McConville, "Uncertainty in GNN Learning Evaluations: A Comparison Between Measures for Quantifying Randomness in GNN Community Detection.", Entropy, 2024. **(Invited Journal)***
- ***Leeney** and McConville, "Unsupervised Optimization of GNNs for Node Clustering.", arXiv, 2024.*

EDUCATION

University of Bristol | 2020-2024

PhD — "**Unsupervised Graph Neural Networks - Training Strategies and Evaluation Principles.**"

- **Proposed novel framework for reliable evaluations.**
- Created metrics for quantifying randomness in experiments.
- Demonstrated **evaluation framework** utility in **privacy preserving federated learning scenario.**
- **Won best talk for Introduction to Neural Networks**
- Unsupervised GNN Frameworks | open-sourced two **PyG** frameworks (UGLE, UGFL) for unsupervised and federated GNN evaluation.
- WRandAi | **PyPI package** for computing the W Randomness Coefficient across ML benchmarks.

University of Oxford | 2021

Machine Learning Summer School

University of Bristol | 2016-2020

MEng, Engineering Mathematics (1st with honours) — "Biologically Inspired **RNN** Hebbian Learning Rule for Decision Making."

PROJECTS

- Plasma Optimization | **differentiable physics** and surrogate models with TORAX to optimise tokamak fusion scenarios — a **9**× fusion-gain improvement.
- Autoresearch | **autonomous research agent** that modifies and evaluates its own training code overnight, keeping changes that improve the target metric.
- Calculator-Augmented LLM | finetuned **Qwen 3.5 2B** to delegate arithmetic via inline computation tokens, with custom **Metal/MLX** backends.
- Earthsea | **AI coding assistant** with a Go server and a Flutter mobile app.
- Weather Wardrobe | **Android** home-screen widget recommending what to wear from the day's weather.
- Storytime | app for listening to **family recordings** of book chapters.
- LeCheck | **computer-vision** pipeline turning sports video into a live team score tracker.
- FPL Agent | **MCP** agent for Fantasy Premier League on AWS Lambda, deployed via **Terraform** and GitHub Actions.
- Money Trees | a leak-free pipeline over **on-chain Bitcoin** transaction-graph data to test for predictive signal.
- Kaei Network Ltd | **Founder** of an AI startup using GNNs to improve academic search; discontinued due to market and funding constraints.
- Obsidian LLM | **TypeScript** interface visualising text generation as a graph.
- Crypto Trader | bespoke **reinforcement learning** trading agent with sentiment data, running on a Raspberry Pi.
- Holistic Spotify Control | gesture control for Spotify using object detection and a **3D CNN**.

TEACHING

I have **supervised over 50 different AI/ML projects for final year undergraduates**, in project creation, technical writing, coding and model evaluation.

- Graphcore | **RNNs** were used to detect cycles in gradient updates to informed adjustments in learning rates.
- Quantum Internet Optimization | **Bayesian Hyperparameter Optimization** was utilized to search for significant input parameters that improved communication bandwidth.
- Anime Auto-Dubber | We applied **OpenAI LLMs** to build a pipeline for converting Japanese audio text. An LTSM NLP model was tuned to translate into English subtitles.
- Campsite Analysis | **CNNs** were used to assess the suitability of land for wild camping from satellite images.
- AI Dance Teacher | **Pose estimation** was combined with custom metrics to measure the quality of dance ability.