Joseph Rance

✉ jr879@cam.ac.uk | 🌐 jr879.user.srcf.net | ⓞ github.com/Joseph-Rance

Research Interests

My interest is in researching more efficient, robust, and trustworthy machine learning algorithms, motivated by a fascination with how complex, intelligent behaviour can emerge using simple learning rules. In particular, I want to understand how we can close the gap in the statistical efficiency and ability to generalise, between the learning rules used by humans and machines, which could significantly improve both AI safety and training costs.

Education

University of Cambridge   2024-2025
MEng in Computer Science
- Dissertation: Constrained Drug Design by Optimising Backward Policies of GFlowNets.
- Supervisors: Miruna Cretu, Pietro Liò
- Modules: Explainable AI (L193), Natural Language Processing (L390), Machine Learning Systems (L46), Reinforcement Learning (R171), Proof Assistants (L81).

University of Cambridge   2021-2024
BA in Computer Science
- First Class in all three years, CST department prize for Highly Commended Part II Dissertation.
- Dissertation: Evaluating attacks on fairness in Federated Learning (link).
- Supervisors: Filip Svoboda, Nicholas Lane.
- Full Blue (fencing)

Colchester Royal Grammar School   2014-2021
Alevels: A*A*A*A*A, GCSEs: 9999999776A*

Publications

Can Private Machine Learning Be Fair? (link)   The 39th Annual AAAI Conference on
Joseph Rance, Filip Svoboda   Artificial Intelligence, 2025

Augmentation Backdoors (link)   ICLR 2023 Workshop on Backdoor Attacks
Joseph Rance, Yiren Zhao, Ilia Shumailov, Robert D. Mullins   and Defenses in Machine Learning

- - - - - - - - - - - -

Evaluating attacks on fairness in federated learning (link)   Dissertation, 2023
Joseph Rance

Experience

Software Engineering Intern, Microsoft   Summer 2023
- Evaluated the performance of the Azure for Operators MLOps codebase under different loads.
- Designed and implemented updates to the MLOps codebase, leading to a 75% cost reduction by processing low-priority data at off-peak times.
- Advocated for a more general framework based on Rust proc macros, leading to my code's integration to the open-source Apache Arrow library (link).
- Presented my work to audiences of more than 30 managers and engineers.
- Led a student team during the Microsoft global hackathon to train a reinforcement learning agent to optimise server scheduling on Azure.

Research Intern, University of Cambridge   Summer 2022
- Developed and tested three new backdoor attacks, which were the first to use compromised data augmentation functions as an attack vector. Our attacks include one of a few existing methods for inserting backdoors with in-distribution data.
- Presented our paper at the ICLR BANDS workshop.
- Supervisors: Yiren Zhao, Ilia Shumailov, Robert Mullins.

Student Volunteer, AI4Good organisation   Summer 2020
- Worked as part of a team to simulate the spread of coronavirus in refugee camps.
- Produced a library of metrics to help evaluate the accuracy of our simulation, which was used to inform decisions made in real camps.

Other Projects

Scaling laws for model pruning based on information theory   2025
- Bounded the information content of two well known computer vision tasks using deep learning.
- Showed that <u>influence functions</u> can efficiently estimate mutual information between a model and its dataset.
- Identified a linear relationship between the size of a pruned model and the information content of its dataset.
- Submitted as undergraduate coursework (combination of <u>L46</u> and <u>L193</u>).

GFlowNets for proof assistants   2025
- Worked with another student to use GFlowNets to prove mathematical theorems in the Lean language.
- Introduced a new state formulation more closely aligned with the task, a new reward function with fewer limitations on generated proofs, and a configuration for fine-tuning with the Mamba architecture.
- Our best model was able to produce correct proofs with a >25% success rate.
- Submitted as undergraduate coursework (<u>R171</u>).

Persistent model tagging using the dying ReLU trick   2025
- Proposed *conditionally-dead* subnetworks - sets of weights that use the dying ReLU problem to force their gradients to 0 - to build backdoors that are resistant to gradient-descent-based unlearning.
- This can be used to 'tag' LLMs by inserting backdoors that prevent the model from posing as a human, while being impossible to remove through conventional finetuning or in-context learning.

Automatic Entrepreneur   2023
- Worked in a team of six students to generate reports on companies based on automatically scraped data.
- Integrated LLMs into the generation pipeline and then used Flask to build an interactive web app.

Robotic arm with object detection   2020
- Led a team of six students to build an autonomous robot arm that used computer vision to identify objects with an onboard camera and pick them up.
- This project was funded by the Jack Petchey Achievement Award.

Using reinforcement learning to evaluate decision making in the sport of fencing   2020
- Developed a set of RL agents to generate tactical policies for the sport of fencing.
- Achieved a 20% improvement in match outcome prediction over the naïve, score-based method.
- I began this project after reading the book *Reinforcement Learning: An Introduction* by Sutton and Barto.

Skills

Languages: Python (TensorFlow, PyTorch), OCaml, Rust, Java, SQL, C/C++, Bash, Prolog, C#, JavaScript, TypeScript, Go, RISC-V assembly, SystemVerilog, LaTeX
Tools: Git, Linux (Ubuntu), Docker, Slurm, Azure, AWS

Awards & Achievements

Awards:
- Robinson College Scholarship
- CST Department Award for Highly Commended Part II Dissertation
- Jack Petchey Achievement Award
- Arkwright Engineering Scholarship
- Cambridge Hawks Award

Competition Results:
- **2nd** UKMT Team Maths Challenge regional finals
- **15th** Aix-en-Provence U20 fencing world cup 2023 (as part of the Belgian team)
- **5th** 2023 Belgian U20 fencing championships
- **1st** 2024 Cambridge Open fencing tournament
- **1st** BUCS Fencing Premier League South (as part of the Cambridge team)