Leo McKee-Reid

AI Safety / ML Researcher

✉ Leo.tzu.mr@gmail.com

📱 +44 0787 4150734

📍 London, UK

PROFESSIONAL SUMMARY

Passionate ML research scientist/engineer leveraging my machine learning skillset to guide the development of safe AI and to reduce existential risk. I am results-oriented and curiosity-driven with a diverse set of experiences in ML, AI safety research (evals, interpretability and field building), computational neuroscience, and aerospace engineering.

EDUCATION

BSc in Computer Science and Mathematics (Internship Distinction) – University of Victoria (2018 – Dec 2023)
- Strategic Framework Experiential Learning Award for computational neuroscience internship.
- Honors status up until final term (dropped out of honors to focus on AI club projects)

Mechanical Engineering Technologist Diploma (Internship Distinction) – Camosun College (2014 - 2016)

WORK EXPERIENCE

Applied ML Researcher – Leap Labs (Oct 2024 – Current)
- Leading the scientific application of an automated ML and AI interpretability engine.
- Wearing many hats on an 8-person lean startup, with a focus on interpretability of various NN architectures.

AI Evals Researcher – LASR Labs (July – Sept 2024)
- Researcher and lead team member on a project evaluating the ability of frontier LLMs to develop deceptive strategies through in-context reinforcement learning.
- Supervised by Mikita Balesni, through the London AI Safety Research (LASR) Labs program.

AI Safety Researcher – AI Safety Camp (Jan – April 2024)
- Worked as team coordinator and researcher on two AI Safety Camp projects: "*Out-of-context learning interpretability*" and "*How promising is automating alignment research?*".

Deep Learning Intern – Tremblay Lab, University of Victoria (Sept 2021 – Sept 2022)
- Built convolutional neural networks with PyTorch to do cell identification and segmentation analysis on volumes of mouse brain image data (from bright-field and electron microscopes).
- Interdisciplinary research with physicists, biologists, and neuroscientists.

AI Safety Researcher – AI Safety Camp (Jan 2022 – June 2022)
- Explored technical solutions to problems surrounding AI deception and human dogmatism within training datasets on a team of aspiring and professional AI safety researchers.

Computational Neuroscience Intern – Salk Institute for Biological Studies (Sept 2019 – April 2020)
- Collaborated on a state-of-the-art neuron and synapse simulation software called MCell at the Computational Neurobiology Lab.
- Interdisciplinary research with physicists, biologists, neuroscientists, and mathematicians.
- Programmed in Python, C, C++, and BioNetGen (a custom language for modeling biochemical reactions).

Production Design Engineer – Rocket Lab (Feb 2017 – Feb 2018)
- Worked in a cutting-edge, face-paced environment with aerospace engineers, machinists, and world class composite technicians to build the world's first orbital rocket with electric-powered turbopumps, 100% carbon fiber fuel tanks, and fully 3D printed engines.
- Designed, procured, and managed the construction of jigs, fixtures, composite molds, large shop structures and composite flight components (notably designed the first system for supporting the entire rocket).
- Coordinated projects to increase build repeatability and decrease manufacturing time and cost of the rocket.

Composite Builder – Rocket Lab (Jan – Feb 2017)
- Conducted research and development with carbon fiber and fiberglass structures.
- Manufactured composite molds and flight components.

Aircraft Design Engineer – Nebula UAS (July – Dec 2016)
- Designed and manufactured complex unmanned aerial vehicles (UAVs) and rapid prototyping.

Team Lead – Center for Aerospace Research (March – Sept 2016)
- Led a team in the design, construction and testing of a custom, high-performance dynamometer.

Jr. Mechanical Engineer – LineSpect (July – Sept 2015)

- Conducted research, construction, and testing of composite airframe manufacturing techniques for autonomous drones.
- Calibration of electronics for flight (ECS, accelerometer, GPS, transmitter, and compass).

ADDITIONAL PROJECTS & EXPERIENCE

AI Alignment Hackathon Participant – Code Red Hackathon, Hosted by METR (March 2024)
- Awarded over $1000 for new ideas and implementing LLM evaluation methods.

AI Club President & Founder – UVic AI, University of Victoria (June 2022 – Dec 2023)
- Hosted technical ML workshops, an AI safety reading group, large hackathons, programming competitions, a public AI safety conference, and led ML projects for the student AI club (uvicai.ca).
- Taught semester-long courses on PyTorch and TensorFlow to fellow undergrads.
- Led a large team of students and gave over 15 workshops and presentations on machine learning and AI safety.
- Raised $11,000 in 2023 for AI and AI safety competitions, conferences, and events.

AlphaZero-like RL Project – UVic AI (Feb – May 2023)
- Led a team that built the first RL model that could play the popular competitive programming game Battlesnake that consistently placed in the top 30 leaderboard during online tournaments.
- Developed a novel modified Monte Carlo tree search algorithm for self-play with a CNN architecture for the 4-player simultaneous move game. Code is currently private as it's continuing to improve for future competitions.
- Won the Most Innovative Project Award at Canada's largest undergrad AI conference (CUCAI).

AI Event Coordinator – Canadian Undergraduate Conference on AI (Nov 2022 – Mar 2023)
- Helped organize sponsors, industry speakers, and presentations at the largest Canadian undergraduate AI conference (350 attendees from 8 different universities) (CUCAI.ca).

Writer/Editor – Martlet Newspaper (Sept 2018 – Aug 2019)
- Conducted interviews, wrote articles, and copy edited for the University of Victoria news organization.

Robotics Club Computer Vision Specialist – UVic Robotics, Student engineering team (Oct 2018 – May 2019)
- Worked on both the Software and Science team in developing an autonomous, Curiosity-like Mars rover to collect and analyze soil samples in the 2019 Canadian International Rover Challenge.

Rocketry Club Team Lead – UVic Rocketry, Student engineering team (Sept 2015 – Jan 2017)
- Led a team of students in the design and manufacturing of 9ft and 12ft carbon fibre rockets.
- Won 3rd place out of 44 rocketry teams from universities around the world in the Intercollegiate Rocket Engineering Competition (IREC 2016) in Utah (successfully launched to 10,000 ft and conducted biological experiments at apogee).

UAV Club Member – UVic AERO, Student engineering team (Sept 2014 – Sept 2016)
- Mechanical engineering member of the design, build, fly UAV/drone club.

PUBLICATIONS

- **Leo M. R.**, Christoph S., Maria A. M., Joe N., Mikita B. Honesty to Subterfuge: In-Context Reinforcement Learning Can Make Honest Models Reward Hack. (2024, NeuIPS SoLaR Workshop).
- **Leo M. R.**, Finn A., Nathan W., Eric S., Dana B. Monte Carlo Tree Search and Reinforcement Learning for a Four Player, Simultaneous Move Game. (2023, CUCAI).
  - In addition, this conference paper and project won the Most Innovative Project award at CUCAI (Canada's largest undergraduate conference for AI).
- Mohammadparsa K., Fernando G. I., Maude B., Katherine P., **Leo M. R.**, Ben-Azu B., Laura M., Marie-Ève T. Manual versus automatic analysis of microglial density and distribution: a comparison in the hippocampus of healthy and lipopolysaccharide-challenged mature male mice. (2022, Micron).

LANGUAGES AND SOFTWARE

Programming Languages and Libraries
- Python (6 years)
- PyTorch (2 years + hosted multiple PT workshops teaching undergrads)
- TensorFlow (1 year + hosted multiple TF workshops teaching undergrads)
- C (5 years)
- C++ (4 years)

Software Tools and Programs
- Blender and CellBlender (computational neuroscience visualization)
- GIMP (writing python scripts for preprocessing image data)
- ImageJ, Ilastik, VAST (neurological imaging visualization and data processing)