

HADRIEN POUGET

✉ pougeth@gmail.com

🔗 hadrienpouget.com

🌐 American and French Dual Citizenship

EDUCATION

University of Oxford Combined Bachelors and Masters in Computer Science 2015 – 2019
1st Class Honours in both 1st and 2nd parts of degree (4.0 GPA as per [Fulbright Commission](#))

WORK EXPERIENCE

Fellow at Center on Long-term Risk *Planned: March 2021 – June 2021*
Awarded £300 for research proposal on investigating the role of work in AI explainability for long-term AI safety.

Research Assistant at University of Oxford *Sep 2019 – Jan 2021*
Working under Daniel Kroening and Thomas Melham, on the verification of Machine Learning systems.

- Designed and implemented Supervised Learning and Reinforcement Learning algorithms in Pytorch and Tensorflow
- Author on four papers investigating robustness of neural networks, and explainability of reinforcement learning
- Attended AI safety reading group, and “Advanced Effective Altruism Fellowship”

Fellow at Sesame Workshop *July 2018 – Sep 2018*
Placement in knowledge and development office,

- Performed data analysis of user behavior in educational apps to draw conclusions about effectiveness.
- Assessed potential for learning analytics within the company and provided recommendations.

Programmer, Kamusi Project Chatbot *July 2017 – Sep 2017*
Coded multi-platform chatbot for Dr. Martin Benjamin (EPFL, Switzerland), for the collection of crowd-sourced translation data.

Extracurricular Activities as a Student *Oct 2015 – June 2019*

- **Teaching assistant** at coding sessions for younger students, **Tutoring**
- **Head of Volunteering for Publicity** at Nightline, a student-run mental health support phone line
- **Peer Supporter** at Oriel College, trained to give mental health support to peers
- **Organization team** for charity event “Jailbreak”, raising nearly £15 000

RESEARCH EXPERIENCE

Academic Papers

- Isaac Dunn, **Hadrien Pouget**, Daniel Kroening and Thomas Melham (2020).
Exposing Previously Undetectable Faults in Deep Neural Networks
 - Under submission to ISSTA
- Isaac Dunn, Laura Hanu*, **Hadrien Pouget***, Daniel Kroening and Thomas Melham (2020).
Evaluating Robustness to Context-Sensitive Feature Perturbations of Different Granularities. [arXiv](#)
 - * equal contribution
 - Accepted to ECCV 2020 workshop “Adversarial Robustness in the Real World”
 - Under submission to IJCAI
 - Presented to Dr. Alessandro Abate’s group in Oxford, Dr. Sanjit Seshia’s group at Berkeley, Dr. Richard Zemel’s group at Toronto and Dr. Alex Aiken’s group at Stanford.

- **Hadrien Pouget**, Hana Chockler, Youcheng Sun and Daniel Kroening (2020). *Ranking Policy Decisions*. [arXiv](#)
 - o Under submission to IJCAI.
- Isaac Dunn, **Hadrien Pouget**, Thomas Melham, and Daniel Kroening (2019). *Adaptive Generation of Unrestricted Adversarial Inputs*. [arXiv](#)

Research Experience as a Student

- **Implementing Reinforcement Learning Techniques on Atari Games – 4th Year Research Project**
1st class honours, *Oct 2018 - June 2019*
Supervised by Dr. Daniel Kroening. Self-taught reinforcement learning, joined reading group, implemented DQN algorithms in tensorflow and pytorch, highlighted inconsistent performance of DQN's by testing readily available implementations.
- **Analysing Variational Continual Learning**
1st class honours, *Feb 2019 – April 2019*
Implemented the algorithm described in *Variational Continual Learning* (Nguyen et al. 2018), bayesian updating of networks. Reproduced and extended results. Worked in a group of 4. Wrote individual reports critically analyzing paper. Made conference style poster presenting findings.
- **Collecting and Analyzing fake news data using Machine Learning – 3rd Year Research Project**
Oct 2017 – June 2018
Supervised by Dr. Varun Kanade. Used webscraping to compile database of news articles and analyzed using different features extractions, models, and training strategies.

ADDITIONAL SKILLS & QUALIFICATIONS

Technical:

Python, Pytorch, Tensorflow, OpenAI Gym, LaTeX, Git

Languages:

Native: English and French

Beginner/intermediate: Spanish

GRE:

Analytical Writing	5/6	92 nd percentile
Quant Reasoning	169/170	95 th percentile
Verbal Reasoning	167/170	98 th percentile

Activities:

Rock climbing, skiing, snowboarding, PADI certified open water diver