Laura **Graves**

Machine Learning Engineer · Researcher

226-581-3641 | lauragravescs@gmail.com | lmgraves | laura-graves-ai

Summary

Machine learning researcher studying at University of Waterloo. Specialization in security and privacy research with experience in bias and fairness research, model testing and verification, and inventing and implementing novel solutions to industry-relevant problems.

Education

University of Waterloo — *Waterloo, ON, Canada*

MASc. in Electrical & Computer Engineering — *Expected Aug. 2021*

- *Proposed Thesis*: Model Patching Through the Amnesiac Machine Learning Framework
- *Relevant Coursework*: Security and Privacy for AI; Reinforcement Learning; Artificial Life Systems, Software Design
- *GPA*: 4.0

Memorial University of Newfoundland — *St. John's, NL, Canada*

B.Sc. in Computer Science, Minor in Mathematics — *April. 2019*

- *GPA*: 3.9 *CS GPA*: 4.0

Preprints & Publications

Amnesiac Machine Learning (in press)

Graves, Laura, V. Nagisetty, V. Ganesh

Proceedings of the AAAI Conference on Artificial Intelligence, 2021

xAI-GAN (in press)

V. Nagisetty, **Graves, Laura,** V. Ganesh

Proceedings of the AAAI workshop on Explainable AI, 2021

Work & Research Experience

Borealis AI — *Toronto, ON, Canada*

Research Intern — *Feb. 2021 - Present*

- Investigating methods for detecting biased or unfair behavior of DNNs
- Developing a novel method for patching models to optimize for a desired fairness metric
- Designing a surrogate feature detection method using gradient alignment and SHAP values
- Comprehensively testing methods on real and synthetic datasets

University of Waterloo — *Waterloo, ON, Canada*

Graduate Researcher — *Sept. 2019 - Present*

- Leading or collaborating on multiple projects across machine learning, software verification, and formal logic fields.
- Led research and development on the Amnesiac Machine Learning project that proposed and evaluated two novel algorithms to selectively remove learned information from trained models (accepted for publication at AAAI-2021).
- Developed architecture, conducted background research, and wrote for the xAI-GAN project, which leverages explainable AI systems to improve GAN training, leading to faster and more stable training (accepted for publication at the AAAI-2021 Explainable AI Workshop).
- Developed fuzzing algorithms, implemented scalable testing methods, and conducted background research for the Constrained Gradient Descent project that tests neural network properties by using differentiable constraints to find property-violating inputs.

Memorial University of Newfoundland — *St. John's, NL, Canada*

Undergraduate Researcher — *Sept. 2018 - Apr. 2019*

- Engineered GAN architecture for learning temporal data through spatial representations as part of the GANs n Reels project, which was featured on CBC radio and television.
- Conducted research on information theory based methods for evaluating information contained in each layer of a DNN.

Skills & Activities

Technical Skills	Machine Learning, Computer Vision, ML Security & Privacy, ML Bias & Fairness, Reinforcement Learning, Model Testing & Verification, Cloud Computing (AWS Sagemaker)
Programming Languages	Python, Java, Javascript, C++
Tools & Frameworks	PyTorch, TensorFlow, Keras, Scikit-Learn, Numpy, Pandas, Jupyter, Matplotlib, Git, REST APIs, Flask, Docker