

JACOB SCHNELL

@ jschnell@uwaterloo.ca 📞 (647) 636-3054 📍 Toronto, ON 🌐 student.cs.uwaterloo.ca/ jschnell/ 📄 schnellj

Experience

Machine Learning Engineering Intern

MAGI Inc.

January – April 2024 (Expected)
New York, NY

- Joining MAGI's machine learning team in the winter.

Lead Research Scientist

Veer Renewables

May 2024 – Present
Vancouver, BC (Remote)

- Leading Veer Renewable's research and development team to develop a diffusion model performing 16x super-resolution on wind data. The model produces high-quality reconstructions of wind patterns 500x cheaper than traditional simulation methods. We are targeting publication.
- Currently focusing on improving the fidelity of reconstructions by leveraging conditional information, such as the terrain elevation, and exploring more sophisticated neural architectures and diffusion process samplers.
- Engineered distributed training and inference pipelines on AWS using PyTorch Distributed, AWS EC2, and Slurm.
- Coordinating Veer's hiring efforts to expand the machine learning team by hiring three new developers. Moving to a management role to guide the team's efforts and guide new hires.

Research Engineering Intern

GPTZero

May – August 2024
Toronto, ON

- Trained and analyzed RoBERTa classifiers, and metric-based classifier to detect AI-written articles with an emphasis on identifying failure cases and robustness to out-of-distribution data. Aided in writing a paper based on these results, currently under review.
- Implemented feature-fusion from Large Language Model (LLM) embeddings and ensembling using metric-based classifiers to decrease the false positive rate in out-of-distribution samples by 21%. Developed four models to fuse LLM embeddings with a RoBERTa classifier.
- Optimized the production pipeline to achieve 13% faster model inference, and developed a parallelized SQL-based dataset resulting in 30% faster model training.

Research Intern

Waabi Innovations

Sep 2023 – Apr 2024
Toronto, ON

- Led a research project to develop a simpler map representation and embedding model for use in autonomous vehicle perception, prediction, and planning under the guidance of Dr. Raquel Urtasun. We are targeting publication.
- The novel representation and transformer-based architecture uses 14% fewer parameters, is 20% faster, and achieves better downstream performance relative to the state-of-the-art LaneGCN architecture.
- Innovated several self-supervised pre-training tasks on vector maps, including masked-autoencoding, contrastive learning, and query classification, improving performance and convergence speed in downstream training.

Undergraduate Research Assistant

University of California Merced

May – Nov 2023
Merced, CA (Remote)

- First author of ScribbleGen, a novel method to generate synthetic training data for use in weakly-supervised semantic segmentation (WSSS). Using a ControlNet diffusion model we generate synthetic images from segmentation scribbles. The resulting image-scribble pairs are used to train a WSSS model.
- Ensured consistency between labels and synthetic images using classifier-free guided diffusion and encode ratios.
- Training using ScribbleGen improves mIoU by 2.3% when training using the full dataset and 6.1% in a low-data setting, achieving a new state-of-the-art for the PascalScribble WSSS dataset.

Research Intern

Waabi Innovations

Jan – May 2023
Toronto, ON

- Developed a novel training scheme for vehicle motion forecasting models under the guidance of Dr. Raquel Urtasun.
- In the new training scheme, motion forecasting is supervised through the trajectory planner to learn predictions that encourage safer driving plans. Training using the new scheme reduced the collision rate by 65%.
- Further experimented with identifying important actors using self-attention weights, to reweight loss by the vehicle's importance, focusing learning efforts on safety-critical vehicles.

Machine Learning Engineering Intern

Flex A.I.

Sep 2021 – Aug 2022

Vancouver, BC (Remote)

- Constructed a novel 4-stage machine learning architecture leveraging pre-trained YOLOv3 and AlphaPose models to identify user errors in workouts from videos. Used this to develop a new data pipeline to automatically process new examples and add them to our dataset.
- Optimized the neural architecture and hyperparameters improving accuracy by 13.4% and training 12x faster compared to our baseline video model (a 3D CNN).
- Developed a multitask image classification model identifying 17 errors across 5 different workout exercises, improving latency and accuracy compared to using individual models.

Publications

- [1] **Schnell, J.**, Wang, J., Qi, L., Hu, V. T., Tang, M., “ScribbleGen: Generative Data Augmentation Improves Scribble-supervised Semantic Segmentation,” in *CVPR 2024 Workshop SyntaGen: Harnessing Generative Models for Synthetic Visual Datasets*, 2024. [Online]. Available: <https://arxiv.org/abs/2311.17121>.
- [2] Adam, G. A., Cui, A., **Schnell, J.**, Thomas, E., Tian, J. J., Dronavalli, A., Tian, E., Lee, D., “GPTZero: Scalable and Explainable Detection of LLM-Generated Texts,” in *submission to EMNLP*, 2024.

Technical Skills

Languages and Tools: Python, C/C++, C#, Java, SQL, Git, \LaTeX , Docker, AWS, Databricks

Libraries: PyTorch, TensorFlow, Keras, PyTorch Lightning, HuggingFace, Scikit-Learn, Numpy, Pandas, Seaborn, Matplotlib

Paradigms: CNNs, RNNs, Transformers, Diffusion Models, GANs, VAEs, Language Modeling, Self-supervised Learning

Education

University of Waterloo

Master of Mathematics in Computer Science

Jan 2024 – Apr 2026

Waterloo, ON

- Supervisors: [Dr. Yuhao Chen](#) and [Dr. Jesse Hoey](#).
- Admitted as part of Waterloo’s Accelerated Master’s Program.

University of Waterloo

Bachelor of Mathematics in Computer Science and Statistics

Sep 2020 – Aug 2025

Waterloo, ON

- 4.0 CGPA, 94% Cumulative average.
- Recipient of the University of Waterloo’s President’s Scholarship of Distinction.

Extracurricular

Conference Reviewer, *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*

2024

Conference Reviewer, *10th Annual Conference on Vision and Intelligent Systems (CVIS)*

2024

Technical Project Manager, *Wat.AI*

July 2024 – Present

- Involved in hiring and managing a team of Waterloo students to collaborate on a technical machine learning project with my partner company Veer Renewables.
- Helping host workshops explaining technical concepts and presenting findings from my project to prospective students.

Reading Group Lead, *University of Waterloo Data Science Club*

Sep 2023 – Present

- I present a machine learning research paper every few weeks to interested university students. Following every presentation I drive a discussion on the applications, limitations, and possible future directions of the paper.
- I help students interested in getting involved in machine learning to advance their careers. This includes providing technical advice for projects and connecting students with potential opportunities.
- My goal is foster a community of passionate and collaborative AI practitioners at the University of Waterloo.