Gaël GENDRON

gael.gendron@auckland.ac.nz ggendro.github.io gael-gendron ggendro scholar

Summary

I am a senior machine learning scientist at LawZero, recipient of the DAAD AInet fellowship on AI for science, publishing in CORE A* AI conferences (EMNLP, ACL, AAAI, IJCAI, AAMAS) and workshops (CaLM@NeurIPS, CRL@NeurIPS, AGI@ICLR). I work on building safe-by-design and trustworthy AI/LLM reasoners that can understand cause-effect relationships and discover new scientific knowledge about the world.

Education

Ph.D **University of Auckland**, Machine Learning, Causality — July 2021 – Sept. 2025
- **Thesis Focus:** Deep Causal Modeling for Reasoning and Generalization

M.Sc. **University of Rennes 1**, Machine Learning for Research — Sept. 2019 – June 2020
- Dual Degree while pursuing a *Diplôme d'Ingénieur*

D. Ing. **National Institute of Applied Science (INSA)**, Software Engineering — Sept. 2018 – June 2020
- *Diplôme d'Ingénieur*, equivalent to M.Sc. Degree

B.Sc. **National Institute of Applied Science (INSA)**, Software Engineering — Sept. 2015 - June 2018

Work Experience

LawZero; Mila - Quebec AI Institute, Senior Machine Learning Scientist — Montréal, CA, Sept. 2025 – Now
- Working on the ScientistAI project aiming at creating safe-by-design AI systems

University of Auckland, Research Assistant — Auckland, NZ, Jan. 2024 – May 2025
- Led the research and development of open-source projects on causal reasoning
- Built the first interpretable neural-causal network for behavior discovery in multi-agent natural systems, with Pytorch, PyG and Lightning

Alten; Amadeus IT Group (contractor), Software Engineer — Sophia Antipolis, FRA, Apr. 2021 – Feb. 2022
- Developed C++ software for a key component of Amadeus' Global Distribution System, enhancing scalability of the retrieval pipeline
- Handled support during code, test and integration phases, management with Scrum
- Managed the project a PO (Project Owner)

University of Auckland, Research Intern — Auckland, NZ, Feb. 2020 – July 2020
- Designed and implemented an open-source project for knowledge graph reasoning using graph neural networks, with Tensorflow

Institute of Electronics and Numerical Technologies, Research Intern — Rennes, FRA, May 2019 – Aug. 2019
- Developed a resource-efficient embedded communication API and reinforcement learning based control module for robotic arms with Keras, Tensorflow, and Dataflow graphs

Inetum, Engineering Intern — Rennes, FRA, July 2018 – Aug. 2018
- Developed a web architecture, SQL database, front-end and back-end of a web platform using NodeJS

Teaching Experience

University of Auckland, Graduate Teaching Assistant — Auckland, NZ, July 2022 – Nov. 2023
- Complexity, Algorithms, Graph Theory, B.Sc. level

- Advanced Machine Learning, M.Sc. level

University of Auckland, PhD Mentor — Jan. 2023 – Jan. 2024

Research Projects

Causal Cartographer — Code

- Built the first end-to-end framework for extraction of real-world causal knowledge and step-by-step counterfactual inference with large language model agents
- Introduced a methodology for provably estimating real-world counterfactuals
- Achieved competitive performance while greatly reducing LLM's context window and output length, decreasing inference cost up to 70% when compared with chain-of-thought

Abstract Reasoning Evaluation — Code

- Conducted the first evaluation of large language models in abstract reasoning and the learning of abstract representations
- Showed that large language models do not extrapolate to unseen reasoning chains

Independent Causal Language Models — Code

- Built and fine-tuned a novel modular language model architecture from causal principles
- Achieved increased efficiency, domain-invariant out-of-distribution reasoning and continual learning without forgetting (increase of o.o.d accuracy up to 40% compared to LoRA fine-tuning)

Latent Space Quantization — Code

- Created a novel variational image auto-encoder based on latent space quantization and causal mechanisms for robust and efficient representation learning

Benchmark Contribution: Humanity's Last Exam — Website

- Contributed to Humanity's Last Exam evaluation benchmark with causal inference tasks

Benchmark Contribution: OpenAI's Evals — Code

- Contributed to OpenAI's LLM evaluation benchmark with abstract reasoning tasks

Honors and Awards

Recipient of the 2024 DAAD AInet Fellowship - AI for Science — Nov. 2024

Invited talk: NAOI Symposium on exploring creativity and intelligence — Sep. 2024

Speaker on the topic of causality and robust reasoning in deep learning

Invited talk: Global Sustainable Development Congress (GSDC) — June 2024

Panelist on the topic of AI, Sustainability and Education with Profs. Siah Hwee Ang (Chair at VUW), Low Teck Seng (Senior Vice President at NUS) and President Banchong Mahaisavariya (Mahidol University)

Recipient of the University of Auckland Best Student Published Paper in Computer Science for "Disentanglement of Latent Representations via Causal Interventions" [Gendron, Witbrock, and Dobbie 2023] — Dec. 2023

Selected Publications

Causal Cartographer: From Mapping to Reasoning Over Counterfactual Worlds **Gaël Gendron**, Joze Rozanec, Michael Witbrock, Gillian Dobbie.

Counterfactual Causal Inference in Natural Language with Large Language Models **Gaël Gendron**, Joze Rozanec, Michael Witbrock, Gillian Dobbie. *Causality and Large Models @NeurIPS 2024*

Robust Domain Generalisation with Causal Invariant Bayesian Neural Networks **Gaël Gendron**, Michael Witbrock, Gillian Dobbie. *Representation Learning @NeurIPS 2024*

Can Large Language Models Learn Independent Causal Mechanisms? **Gaël Gendron**, Bao Trung Nguyen, Alex

Peng, Michael Witbrock, Gillian Dobbie. *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, EMNLP 2024*

Exploring iterative enhancement for improving learnersourced multiple-choice question explanations with large language models Qiming Bao, Juho Leinonen, Alex Peng, Wanjun Zhong, **Gaël Gendron**, et al. *Proceedings of the 39th AAAI Conference on Artificial Intelligence (2025)*

Large Language Models Are Not Strong Abstract Reasoners **Gaël Gendron**, Qiming Bao, Michael Witbrock, Gillian Dobbie. *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI 2024 - How Far Are We From AGI? @ICLR 2024*

Causal Graph Modeling with Deep Neural Engines for Strong Abstract Reasoning in Language and Vision **Gaël Gendron**. *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI 2024*

Abstract Meaning Representation-Based Logic-Driven Data Augmentation for Logical Reasoning Qiming Bao, Alex Yuxuan Peng, Zhenyun Deng, Wanjun Zhong, **Gaël Gendron**, et al. *Findings of the Association for Computational Linguistics, ACL 2024*

Behaviour Modelling of Social Animals via Causal Structure Discovery and Graph Neural Networks **Gaël Gendron**, Yang Chen, Mitchell Rogers Yiping Liu, Mihailo Azhar, Shahrokh Heidari, et al. *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2024*

Meerkat Behaviour Recognition Dataset Mitchell Rogers, **Gaël Gendron**, David Arturo Soriano Valdez, Mihailo Azhar, Yang Chen, et al. *Computer Vision for Animal Behavior Tracking and Modeling @CVPR 2023*

Disentanglement of Latent Representations via Causal Interventions **Gaël Gendron**, Michael Witbrock, Gillian Dobbie. *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI 2023*

Reviewing

Program Committee for the Fortieth AAAI Conference on Artificial Intelligence (AAAI 2026)	Aug. 2025
Reviewer for the Thirty-Ninth Annual Conference on Neural Information Processing Systems (NeurIPS 2025)	Jun. 2025
Reviewer for the Second Conference on Language Modeling (COLM 2025)	Apr. 2025
Reviewer for the Thirteenth International Conference on Learning Representations (ICLR 2025)	Oct. 2024
Reviewer for the Computer Vision for Animal Behavior Tracking and Modeling Workshop at the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2024 (CV4Animals@CVPR 2024)	Apr. 2024
Emergency Reviewer for the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2023)	Aug. 2023
External Reviewer for the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Database (ECML PKDD 2023)	Apr. 2023

Skills

Programming Languages: Python, C++, Java, Javascript

Machine Learning: Probabilistic Graphical Causal Models, Causal Inference, Computer Vision, Natural Language Processing, Large Language Models, Prompt-Tuning, Fine-Tuning, Post-Training, Graph Neural Networks, Reinforcement Learning, Variational Autoencoders, Bayesian Neural Networks, OOD Generalization, AI Safety, AI For Science

Libraries: Pytorch, Pytorch Lightning, Pytorch Geometric, Tensorflow, Scikitlearn, HuggingFace Transformers, Langchain, Smolagents, DoWhy, InspectAI, NLTK

Languages: French (native speaker), English (proficient/fluent/C1), Spanish (intermediate/B1), Chinese (basic,A2)