

ANDREW JESSON

Brooklyn, NY

andrew.d.jesson@gmail.com • andrewjesson.com

AI researcher building self-improving LLM systems and the infrastructure to evaluate them. Ten years of research with first-author publications at NeurIPS, ICML, and ICLR (4,700+ citations); founding technical staff at TensorZero — core contributor to its open-source LLM stack (11,000+ GitHub stars) and lead developer of LLM Gym.

RESEARCH INTERESTS

Agent Self-Improvement and Harness Engineering, Reinforcement Learning and Post-Training, Uncertainty Quantification for Generative AI, Causal Inference and Experimentation

CURRENT POSITION

TensorZero, New York

Jul 2025 – Present

Founding Member of Technical Staff – Research

- Building [TensorZero](#), an open-source LLM infrastructure stack unifying a high-performance gateway, observability, optimization, evaluations, and experimentation (11,000+ GitHub stars; \$7.3M seed, August 2025); core contributor (#6 of 118) with 500+ merged PRs across the TensorZero organization.
- Research on automated agent engineering: [the engineering practices coding agents like Claude Code and Codex use to improve AI agents](#) and [when data helps automated agent engineering](#).
- Research on LLM post-training and optimization: [distillation with programmatic data curation](#) (matching frontier-model quality at 5–30x lower inference cost; [front page of Hacker News](#)), [evaluating OpenAI’s Reinforcement Fine-Tuning](#), and [scaling automated prompt engineering to agentic tasks](#).
- Building [LLM Gym](#) (lead developer), a unified environment interface for developing and benchmarking LLM applications that learn from feedback (agentic tool use, software engineering, RAG, games, and more); it powers internal research and development, including experimental results for the above research, [TensorZero Autopilot, an automated AI engineer](#), and [analysis of noisy LLM evaluators](#).

PROFESSIONAL EXPERIENCE

TensorZero, New York

Jan 2025 – Jun 2025

Research Consultant

- Initiated [LLM Gym](#) while completing postdoctoral research at Columbia.

Columbia University, New York

Jul 2023 – Jun 2025

Postdoctoral Researcher

- Department of Statistics; advised by Professors David Blei and Donald Green.
- Uncertainty quantification for generative AI: estimating hallucination rates (NeurIPS 2024) and predicting when in-context learning succeeds (ICLR 2025).

Meta, London

Jun 2022 – Sep 2022

Intern – Research Scientist

- Human guided automated content moderation

Imagia, Montreal

Sep 2017 – Sep 2019

Program Manager

- Radiomics, Robustness, Generative Models.

- Imagia, Montreal** Sep 2016 – Aug 2017
Applied Research Scientist
- Deep learning algorithms for detection of lung nodules in CT images.
 - GAN based unsupervised clustering of high dimensional data.
- Imagia, Montreal** May 2016 – Aug 2016
Intern – NSERC CREATE-MIA
- Deep learning algorithms for segmentation in medical imaging.
- NeuroRx, Montreal** May 2014 – Aug 2014
Intern – NSERC CREATE-MIA
- Evaluation software for MS Lesion detection algorithms.
- CMC Electronics, Montreal** Sep 2013 – Dec 2013
Intern – Full-stack developer
- Web applications.

EDUCATION

- University of Oxford – PhD** Oct 2019 – Jun 2024
 Oxford Applied and Theoretical Machine Learning Group (OATML)
 Department of Computer Science
 Advised by Yarín Gal
 Thesis: *Structural and Statistical Uncertainty in Observational Causal Machine Learning at Scale*
- McGill University, Montreal – Master’s Degree** Sep 2015 – Jun 2018
 Probabilistic Vision Group (PVG)
 Department of Electrical and Computer Engineering
 Advised by Tal Arbel
- McGill University, Montreal – Bachelor of Engineering** Sep 2011 – Jun 2015
 Department of Electrical and Computer Engineering
- Concordia University, Montreal – Bachelor of Arts** Sep 2007 – Jun 2011
 Major in Philosophy

SELECTED PUBLICATIONS

- 4,700+ citations, h-index 20 – [Google Scholar profile](#)
- Andrew Jesson**, Nicolas Beltran-Velez, David Blei. [Can Generative AI Solve Your In-Context Learning Problem? A Martingale Perspective](#). ICLR. 2025.
- Andrew Jesson**, Nicolas Beltran-Velez, Quentin Chu, Sweta Karlekar, Jannik Kossen, Yarín Gal, John P. Cunningham, David Blei. [Estimating the Hallucination Rate of Generative AI](#). NeurIPS. 2024. (50+ citations)
- Claudia Shi, Nicolas Beltran-Velez, Achille Nazaret, Carolina Zheng, Adrià Garriga-Alonso, **Andrew Jesson**, Maggie Makar, David Blei. [Hypothesis Testing the Circuit Hypothesis in LLMs](#). NeurIPS. 2024. Oral and 3rd Prize, ICML 2024 Mechanistic Interpretability Workshop.
- Andrew Jesson**, YiDing Jiang. [Improving Generalization on the ProcGen Benchmark with Simple Architectural Changes and Scale](#). arXiv preprint arXiv:2410.10905. 2024.
- Andrew Jesson**, Chris Lu, Gunshi Gupta, Angelos Filos, Jakob Nicolaus Foerster, Yarín Gal. [ReLU to the Rescue: Improve Your On-Policy Actor-Critic with Positive Advantages](#). ICML. 2024. <https://github.com/anndvision/vsop>

Miruna Oprescu, Jacob Dorn, Marah Ghoummaid, **Andrew Jesson**, Nathan Kallus, Uri Shalit. [B-Learner: Quasi-Oracle Bounds on Heterogeneous Causal Effects Under Hidden Confounding](#). ICML. 2023.

Yashas Annadani, Panagiotis Tigas, Desi R Ivanova, **Andrew Jesson**, Yarin Gal, Adam Foster, Stefan Bauer. [Differentiable Multi-Target Causal Bayesian Experimental Design](#). ICML. 2023.

Clare Lyle, Arash Mehrjou, Pascal Notin, **Andrew Jesson**, Stefan Bauer, Yarin Gal, Patrick Schwab. [DiscoBAX: Discovery of optimal intervention sets in genomic experiment design](#). ICML. 2023.

Shreshth A Malik, Salem Lahlou, **Andrew Jesson**, Moksh Jain, Nikolay Malkin, Tristan Deleu, Yoshua Bengio, Yarin Gal. [Batchgfn: Generative flow networks for batch active learning](#). ICML workshop on Structured Probabilistic Inference & Generative Modeling. 2023.

Myrl G Marmarelis, Elizabeth Haddad, **Andrew Jesson**, Neda Jahanshad, Aram Galstyan, Greg Ver Steeg. [Partial identification of dose responses with hidden confounders](#). UAI. 2023.

Andreas Kirsch, Sebastian Farquhar, Parmida Atighehchian, **Andrew Jesson**, Frédéric Branchaud-Charron, Yarin Gal. [Stochastic Batch Acquisition: A Simple Baseline for Deep Active Learning](#). TMLR. 2023. (60+ citations)

Andrew Jesson, Alyson Douglas, Peter Manshausen, Maëlys Solal, Nicolai Meinshausen, Philip Stier, Yarin Gal, Uri Shalit. [Scalable Sensitivity and Uncertainty Analysis for Causal-Effect Estimates of Continuous-Valued Interventions](#). NeurIPS. 2022. <https://github.com/anndvision/overcast>

Panagiotis Tigas*, Yashas Annadani*, **Andrew Jesson**, Bernhard Schölkopf, Yarin Gal, Stefan Bauer. [Interventions, Where and How? Experimental Design for Causal Models at Scale](#). NeurIPS. 2022. (75+ citations)

Maëlys Solal, **Andrew Jesson**, Yarin Gal, Alyson Douglas. [Using uncertainty-aware machine learning models to study aerosol-cloud interactions](#). Tackling Climate Change with Machine Learning: workshop at NeurIPS. 2022.

Arash Mehrjou, Ashkan Soleymani, **Andrew Jesson**, Pascal Notin, Yarin Gal, Stefan Bauer, Patrick Schwab. [GeneDisco: A Benchmark for Experimental Design in Drug Discovery](#). ICLR. 2022.

Andrew Jesson*, Panagiotis Tigas*, Joost van Amersfoort, Andreas Kirsch, Uri Shalit, Yarin Gal. [Causal-BALD: Deep Bayesian Active Learning of Outcomes to Infer Treatment-Effects from Observational Data](#). NeurIPS. 2021. <https://github.com/OATML/causal-bald>

Andrew Jesson*, Peter Manshausen*, Alyson Douglas*, Duncan Watson-Parris, Yarin Gal, Philip Stier. [Using Non-Linear Causal Models to Study Aerosol-Cloud Interactions in the Southeast Pacific](#). Causal Inference & Machine Learning: Why now? Workshop, NeurIPS. 2021. <https://github.com/anndvision/clouds>

Joost van Amersfoort, Lewis Smith, **Andrew Jesson**, Oscar Key, Yarin Gal. [On Feature Collapse and Deep Kernel Learning for Single Forward Pass Uncertainty](#). Bayesian Deep Learning workshop, NeurIPS 2021. (220+ citations) <https://github.com/y0ast/DUE>

Andrew Jesson, Sören Mindermann, Yarin Gal, Uri Shalit. [Quantifying Ignorance in Individual-Level Causal-Effect Estimates under Hidden Confounding](#). ICML. 2021. (100+ citations) <https://github.com/anndvision/quince>

Andrew Jesson*, Sören Mindermann*, Uri Shalit, Yarin Gal. [Identifying Causal-Effect Inference Failure with Uncertainty-Aware Models](#). NeurIPS. 2020. (110+ citations) <https://github.com/OATML/ucate>

Aaron Carass, et al. (including **Andrew Jesson**). [Evaluating White Matter Lesion Segmentations with Refined Sørensen-Dice Analysis](#). Scientific Reports. 2020. (260+ citations)

Andrew Jesson, Cécile Low-Kam, Tanya Nair, Florian Soudan, Florent Chandelier, Nicolas Chapados [Adversarially Learned Mixture Model](#). Theoretical Foundations and Applications of Deep Generative Models Workshop, ICML. 2018.

Spyridon Bakas, et al. (including **Andrew Jesson**). [Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge](#). arXiv preprint arXiv:1811.02629. 2018. (2,800+ citations)

Andrew Jesson, Nicolas Guizard, Sina Hamidi Ghalehjegh, Damien Goblot, Florian Soudan, Nicolas Chapados [CASED: Curriculum Adaptive Sampling for Extreme Data Imbalance](#). MICCAI. 2017. (60+ citations)

Aaron Carass, et al. (including **Andrew Jesson**). [Longitudinal multiple sclerosis lesion segmentation: Resource and challenge](#). NeuroImage. 2017. (460+ citations)

Andrew Jesson, Tal Arbel. [Brain tumor segmentation using a 3D FCN with multi-scale loss](#). Brain Lesion Workshop, MICCAI. 2017.

Andrew Jesson, Tal Arbel. [Hierarchical MRF and random forest segmentation of MS lesions and healthy tissues in brain MRI](#). Longitudinal Multiple Sclerosis Lesion Segmentation Challenge, ISBI. 2015.

INVITED TALKS

- | | |
|--|----------|
| Andrew Gordon Wilson's Group, NYU – <i>Can Generative AI Solve Your In-Context Learning Problem?</i> | Oct 2024 |
| Machine Learning and Friends Lunch, UMass Amherst – Causal Inference | Sep 2023 |
| RISE Research Institutes of Sweden – <i>Head in the Clouds? Why Decision Making is Hard</i> | 2022 |
| Vilnius Machine Learning Workshop (EEML) – <i>Identifying Causal-Effect Inference Failure with Uncertainty-Aware Models</i> | 2021 |

ACADEMIC SERVICE

- | | |
|---|-------------|
| Program Committee , Machine Learning for Drug Discovery (MLDD) Workshop, ICLR | 2023 |
| Co-organizer , Machine Learning for Drug Discovery (MLDD) Workshop, ICLR | 2022 |
| Organizer and Host , OATML Invited Talks Series , University of Oxford | 2019 – 2020 |

SUPERVISION

Maëlys Solal

Dissertation: *Machine Learning Approaches For Estimating The Causal Effects Of Aerosols On Clouds*.
MSc in Advanced Computer Science, University of Oxford, 2022.
Co-supervised with Alyson Douglas and Yarin Gal

Daniel Ritter

Dissertation: *Assessing the Interpretability of Large Language Models*.
MSc in Advanced Computer Science, University of Oxford, 2022.
Co-supervised with Lisa Schüt, Yarin Gal, and Been Kim

ACADEMIC POSITIONS

- | | |
|--|---------------------|
| McGill University, Montreal
<i>Teaching Assistant</i> | Sep 2015 – Apr 2016 |
| <ul style="list-style-type: none">• Introduction to Computer Engineering. | |
| McGill University, Montreal
<i>Research Assistant</i> | May 2015 – Aug 2015 |
| <ul style="list-style-type: none">• Probabilistic methods for detection of MS lesions. | |

COMPETITIONS

MICCAI Multimodal Brain Tumor Segmentation Challenge (BraTS) 2017	Top 10
ISBI 2015 Longitudinal Multiple Sclerosis Lesion Segmentation Challenge	2nd Place

AWARDS AND SCHOLARSHIPS

3rd Prize, Mechanistic Interpretability Workshop, ICML	2024
Russel Studentship (£170,000)	2019–2023
NSERC CREATE in Medical Image Analysis (\$13,500)	2016
NSERC CREATE in Medical Image Analysis (\$13,500)	2015
Summer Undergraduate Researcher in Engineering (SURE) (\$6,000)	2014
NSERC CREATE in Medical Image Analysis Fellow	2014

PATENTS

Florent Chandelier, Andrew Jesson, Mohammad Havaei, Lisa Dijorio, Cécile Low-Kam, Nicolas Chapados. [Method and system for generating synthetically anonymized data for a given task](#). US Patent App. 17/259,908. 2021.