Elliot Creager

University of Waterloo – 200 University Ave – Waterloo, ON N2L 3G3
creager@uwaterloo.ca • ecreager.github.io
Google Scholar: boeblUcAAAAJ • Semantic Scholar: 3422145

Current Position

Assistant Professor (Department of Electrical and Computer Engineering) — University of Waterloo
2023 – Present — Waterloo, ON, Canada

- Director, Socially Embedded Machine Intelligence (SEMI) Lab
- Cross-appointed Faculty, David R. Cheriton School of Computer Science
- Cross-appointed Faculty, Department of Management Science and Engineering
- Faculty Affiliate, Vector Institute for Artificial Intelligence
- Faculty Affiliate, Schwartz Reisman Institute for Technology and Society
- Faculty Affiliate, Data and AI Institute at University of Waterloo
- Faculty Affiliate, Cybersecurity and Privacy Institute at University of Waterloo
- Faculty Affiliate, AI Ready Lab at University of Waterloo
- Faculty Affiliate, Digital Democracies Institute at Simon Fraser University

Research Interests

Trustworthy and Responsible AI · User Data Autonomy · AI Safety
Algorithmic Fairness · Privacy-preserving Machine Learning · Deep Learning

Preprints

† denotes equal contribution

[Fal+26] Ali Falahati, **Elliot Creager**, Gautam Kamath, and Shubhankar Mohapatra. "DriftXpress: Faster Drifting Models via Projected RKHS Fields". 2026.

Conference Papers (Last 6 years)

† denotes equal contribution

Note: In machine learning, conference proceedings are the main venue for publishing complete research works. Workshops are used to publish smaller projects and works in progress.

[Ham+26] Mysha Hamid, Asmi Gujral, Iman Umair-Qaiser, Emma Pileggi, Chelsea Dmytryk, Sameek Sharma, Jady Xu, Elbert Chen, Nikola Jokić, Megan Chang, and **Elliot Creager**. "Investigating Reinforcement Learning for Adaptive Signal Transmission in High-Altitude Communication". In: *77th International Astronautical Congress (IAC), IAF Space Communications and Navigation Symposium, Interactive Presentations*. Accepted abstract, Paper ID 113682. 2026.

[BAC26] Meghana Bhange, Ulrich Aïvodji, and **Elliot Creager**. "Test-Time Collective Action: Proxy-Based Perturbations for Correcting Algorithmic Harms". In: *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES)*. 2026.

[Tri+26] Huy Trinh, Phuong Thai, **Elliot Creager**, and George Shaker. "Reliable Quasi-Static Post-Fall Floor-Occupancy Detection Using Low-Cost Millimetre-Wave Radar". In: *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2026.

[Cha+26] Cléa Chataigner†, Rebecca Ma†, Prakhar Ganesh, Afaf Taïk, **Elliot Creager**, and Golnoosh Farnadi. "Say It Another Way: Auditing LLMs with a User-Grounded Automated Paraphrasing Framework". In: *Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*. **Oral (top 14%)**. 2026.

[Tri+25] Huy Trinh, Sebastian Ratto V., **Elliot Creager**, and George Shaker. "A Physics-Informed Digital Twin Framework for Calibrated Sim-to-Real FMCW Radar Occupancy Estimation". In: *University of Waterloo Conference on Vision and Intelligent Systems (CVIS)*. **Oral**. 2025.

[Sol+26] Rushabh Solanki, Meghana Bhange, Ulrich Aïvodji, and **Elliot Creager**. "Crowding Out The Noise: Algorithmic Collective Action Under Differential Privacy". In: *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAccT)*. 2026.

[Sol+25] Rushabh Solanki, Meghana Bhange, Ulrich Aïvodji, and **Elliot Creager**. "Algorithmic Collective Action Directed at Fair and Private Learning Algorithms". In: *University of Waterloo Cybersecurity and Privacy Institute Graduate Student Conference*. **Spotlight**. 2025.

[Ala+24] Parand A. Alamdari, Toryn Q. Klassen, **Elliot Creager**, and Sheila A. McIlraith. "Remembering to be Fair: Non-Markovian Fairness in Sequential Decision Making". In: *International Conference on Machine Learning (ICML)*. PMLR. 2024, pp. 906–920.

[Eyr+24] Benjamin Eyre, **Elliot Creager**, David Madras, Vardan Papyan, and Richard Zemel. "Out of the Ordinary: Spectrally Adapting Regression for Covariate Shift". In: *International Conference on Machine Learning (ICML)*. PMLR. 2024, pp. 12701–12722.

[Man+23] Arjun Mani, Ishaan Preetam Chandratreya, **Elliot Creager**, Carl Vondrick, and Richard Zemel. "SURFSUP: Learning Fluid Simulation for Novel Surfaces". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023, pp. 14225–14235.

[Pit+22] Silviu Pitis, **Elliot Creager**, Ajay Mandlekar, and Animesh Garg. "MoCoDA: Model-based Counterfactual Data Augmentation". In: *Neural Information Processing Systems (NeurIPS)*. Vol. 35. 2022, pp. 18143–18156.

[Trä+21] Frederik Träuble, **Elliot Creager**, Niki Kilbertus, Francesco Locatello, Andrea Dittadi, Anirudh Goyal, Bernhard Schölkopf, and Stefan Bauer. "On Disentangled Representations Learned from Correlated Data". In: *International Conference on Machine Learning (ICML)*. **Oral (top 3%)**. PMLR. 2021, pp. 10401–10412.

[CJZ21] **Elliot Creager**, Jörn-Henrik Jacobsen, and Richard Zemel. "Environment Inference for Invariant Learning". In: *International Conference on Machine Learning (ICML)*. PMLR. 2021, pp. 2189–2200.

Journal Articles

† denotes equal contribution

[TCS26] Huy Trinh, **Elliot Creager**, and George Shaker. "Doppler-Domain Warping and Resampling for Semi-Static Human Occupancy Detection Using Low-Resolution SIMO FMCW Radar". In: *IEEE Transactions on Radar Systems* (2026).

Workshop Papers (Peer reviewed; Last 6 years)

† denotes equal contribution

[HEC25] Joshua Hill, Ben Eyre, and **Elliot Creager**. Transformers Don't In-Context Learn Least Squares Regression. ICML 2025 Workshop on Reliable and Responsible Foundation Models. **Oral (top 5%)**. 2025.

[Zha+25] Minhui Zhang, Prahar Ijner, Yoav Wald, and **Elliot Creager**. Active Slice Discovery in Large Language Models. NeurIPS Workshop on Reliable ML from Unreliable Data. 2025.

[Lib+25] Lena Libon, Meghana Bhange, Rushabh Solanki, **Elliot Creager**, and Ulrich Aïvodji. Conscious Data Contribution via Community-Driven Chain-of-Thought Distillation. NeurIPS Workshop on Algorithmic Collective Action. **Oral (top 10%)**. 2025.

[Ngu+25] Chloe Nguyen, Jin Sol Kim, Lai-Tze Fan, and **Elliot Creager**. Out of Sight, Out of Mind: Problems Perpetuated by Training on Facial Images Collected by De Facto Consent. IEEE ETHICS 2025 Undergraduate Student Track. 2025.

[SC24] Rushabh Solanki and **Elliot Creager**. Promoting User Data Autonomy During the Dissolution of a Monopolistic Firm. NeurIPS Workshop on Regulatable ML. 2024.

[Jai+24] Brandon Jaipersaud, Zining Zhu, Frank Rudzicz, and **Elliot Creager**. Show, Don't Tell: Uncovering Implicit Character Portrayal using LLMs. NeurIPS Workshop on Creativity and Generative AI. 2024.

[EZC22] Benjamin Eyre, Richard Zemel, and **Elliot Creager**. Towards Environment-Invariant Representation Learning for Robust Task Transfer. ICML 2022 Workshop on Spurious Correlations, Invariance and Stability. 2022.

[DC21] Dilys Dickson and **Elliot Creager**. Measuring User Recourse in a Dynamic Recommender System. ICML 2021 Workshop on Algorithmic Recourse. 2021.

[CZ21] **Elliot Creager** and Richard Zemel. Online algorithmic recourse by collective action. ICML 2021 Workshop on Algorithmic Recourse. 2021.

Selected Talks & Presentations

Montreal Auditing Workshop (Type: Invited Talk) June 23, 2026
Title: Attribution Multiplicity: Understanding Uncertainty in Data Valuations Montréal, QC, Canada

MSE 720 (Responsible AI) (Type: Guest Lecture) June 16, 2026
Title: The Role of Coordination and Collective Action in Trustworthy ML Waterloo, ON, Canada

Canadian AI Conference, Workshop on Data and Model Protection in GenAI
(Type: Invited Talk) May 25, 2026
Title: The Role of Coordination and Collective Action in Trustworthy ML Vancouver, BC, Canada

Western University (Starling Centre) (Type: Invited Talk) April 22, 2026
Title: The Role of Coordination and Collective Action in Trustworthy ML London, ON, Canada

AI Graduate Seminar (University of Waterloo CS) (Type: Invited Talk) February 26, 2026
Title: The Role of Coordination and Collective Action in Trustworthy ML Waterloo, ON, Canada

Youth Tech Labs (AI Education for High Schoolers) (Type: Keynote) October 18, 2025
Title: How Machine Learning Turns (Your) Data Into Predictive Models Waterloo, ON, Canada

ACM/SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)
(Type: Invited Panelist) August 04, 2025
Title: Responsible AI: Challenges and Opportunities Toronto, ON, Canada
Notes: Responsible AI Special Day panel

ICML Workshop on Reliable & Responsible Foundation Models (Type: Contributed Talk) July 19, 2025
Title: Transformers Don't In-context Learn Least Squares Regression Vancouver, BC, Canada
Notes: Oral presentation (**top 5%** of accepted papers)

Huawei Workshop on Conversational User Interfaces (Type: Invited Talk) July 11, 2025
Title: Stress Testing the Building Blocks of Conversational AI Waterloo, ON, Canada

Waterloo.AI Seminar Series (Type: Invited Talk) June 13, 2024
Title: To Build AI That Works For Everyone, Adapt Models and Coordinate Data Waterloo, ON, Canada

ECE Seminar Series (Type: Invited Talk) June 04, 2024
Title: Building Fair and Reliable AI Systems Waterloo, ON, Canada

EECS 6320: Fairness and Bias in AI (York University) (Type: Guest Lecture) March 15, 2024
Title: Fairness and Robustness in Representation Learning Toronto, ON, Canada

CSC 2421: Foundations of Algorithmic Fairness (U. of Toronto) (Type: Guest Lecture) March 14, 2024
Title: Algorithmic Fairness: from Representation Learning to Robustness Toronto, ON, Canada

EQUAL Lab Meeting (McGill University) (Type: Invited Talk) November 13, 2023
Title: Out of the Ordinary: Spectrally Adapting Regression for Covariate Shift Montréal, QC, Canada

Causal Inference for Improved Policy Workshop (UofT DSI) (Type: Invited Talk) November 10, 2023
Title: Methods for Counterfactual Data Augmentation in Reinforcement Learning Toronto, ON, Canada

ML Privacy and Security Workshop (Vector Institute) (Type: Invited Talk) July 10, 2023
Title: Can "Adversaries" Play a Positive Role in Ethical AI? Toronto, ON, Canada

DEFirst Reading Group (Mila) (Type: Invited Talk) April 27, 2023
Title: Methods for Counterfactual Data Augmentation in Reinforcement Learning Montréal, QC, Canada

Pursue STEM Outreach Program for High Schoolers (Type: Invited Talk) February 23, 2023
Title: Society and Ethics Concerns in Machine Learning Toronto, ON, Canada

NeurIPS Highlights (Vector Institute) (Type: Invited Talk) November 17, 2022
Title: Methods for Counterfactual Data Augmentation in RL Toronto, ON, Canada

Bias in AI for Industry Partners (Vector Institute) (Type: Invited Talk) March 29, 2022
Title: Bias in AI: Mitigation Strategies Remote

Domain Adaptation and Transfer Workshop (MICCAI 2021) (Type: Keynote) October 21, 2021
Title: A Fairness Perspective on Robust Representation Learning Remote

International Conference on Machine Learning (Type: Contributed Talk) July 10, 2021
Title: Environment Inference for Invariant Learning Remote

Endless Summer School (Vector Institute) (Type: Invited Talk) April 21, 2021
Title: Fair Representation Learning with Disentanglement Remote

Bias in AI for Industry Partners (Vector Institute) (Type: Invited Talk) February 22, 2021
Title: Bias in AI: Mitigation Strategies Remote

Bias in AI Meeting (Princeton University) (Type: Invited Talk) March 06, 2020
Title: Causal Modeling for Fairness in Dynamical Systems Princeton, NJ, USA

Research Funding

INOVAIT Pilot Fund
2026 – 2027
Title: Triaging AI tool for pulmonary health using Spectral X-ray imaging technology

Mellon Higher Learning Program
2026 – 2028
Title: Data Fluencies 2.0: Understanding AI (Waterloo Subgrant)

CIFAR/CAISI Catalyst Grant
2026 – 2028
Title: Democratic Alignment of LLMs Through Economic Theory: Relative Preferences and Strategic Coordination

University of Texas Global Research Seed Grant
2025 – 2026
Title: Fostering Sustainable and Equitable Worker-AI Collaboration through Algorithmic Collective Action

SSHRC Insight Grant
2025 – 2028
Title: Interdisciplinary Approaches toward Responsible Facial Recognition AI: Developing Technical, Ethical, and Regulatory Recommendations for Policymaking in Canada

Waterloo Data and AI Institute Seed Fund
2024 – 2025
Title: Can Facial Recognition Technologies be both Racially Diverse and Ethical?

CIFAR AI Catalyst
2024 – 2026
Title: AI Auditing through Exploration of Model Multiplicity

NSERC Discovery Grant
2024 – 2028
Title: Responsible and Equitable AI: Tools for Model Owners and Data Subjects

University of Waterloo New Faculty Startup Award
2023 – Present

Teaching

ECE 750T: Topics in Computer Software: Responsible AI Fall 2026
University of Waterloo

ECE 457B: Machine Learning Fall 2026
University of Waterloo

ECE 493 / ECE 657D: Neural Networks and Deep Learning Winter 2026
University of Waterloo

ECE 406: Algorithm Design and Analysis Winter 2024, Winter 2025
University of Waterloo

CSC 384: Introduction to Artificial Intelligence Winter 2022
University of Toronto

Student Supervision

Postdoctoral Supervision

Jin Sol Kim (Postdoc)
2025 – Present
Project: How do the AI Ethics of Facial Recognition inform Canadian Policy?

Doctoral Supervision

Soudeh Nilforoushan (PhD, ECE)
2026 – Present
Thesis: Data Attribution for Trustworthy Machine Learning

Meghana Bhange (PhD, ÉTS Montréal/Mila)
2025 – Present
Thesis: Repair from Below by Algorithmic Collective Action

Master's Supervision

Ambreesh Parthasarathy (MASc, ECE)
2025 – Present
Thesis: Using Synthetic Data for Fair Diffusion Causes Tokenization Harms
Awards: International Master's Award of Excellence

Brandon Jaipersaud (MASc, ECE)
2025 – Present
Thesis: Exploring Alignment Faking in Large Language Models through Data Attribution
Awards: CGS-M; OGS-M (declined)

Huy Trinh (MASc, ECE)
2024 – 2026
Thesis: Efficient and Scalable AI at the Edge to Support Healthy Aging

Rebecca Ma (MASc, ECE)
2024 – Present

Thesis: Robust Auditing of Large Language Models

Rushabh Solanki (MASc, ECE)
2024 – 2025
Thesis: Algorithmic Collective Action in Trustworthy AI Systems

Dilys Dickson (MMI, African Institute for Mathematical Sciences)
2020 – 2021
Thesis: Measuring User Recourse in a Dynamic Recommender System

Undergraduate Supervision

Ahmed Osman (Ugrad Co-op, ECE)
2026 – 2026
Project: Exploring Transfer Learning in Medical Foundation Models

Deacon Sham (Ugrad URA, ECE)
2026 – 2026
Project: Building a RAG System for Facial Recognition Policymakers

Catherine Jin (Ugrad Co-op, CS)
2026 – 2026
Project: The Data Cartographer: Visualizing Influence of Training Data in LLMs

Alisha Anat (Ugrad Co-op, ECE)
2026 – 2026
Project: AI Primary Source Retrieval for Policy Makers

Yuhao Chen (Ugrad Co-op, CS)
2025 – 2025
Project: A Representation Geometry Perspective on LLM Prompt Sensitivity

Chloe Nguyen (Ugrad Co-op, SYDE)
2025 – 2025
Project: Problems Perpetuated by Training on Facial Images Collected by De Facto Consent

Mysha Hamid, Emma Pileggi, Sameek Sharma, Asmi Gujral (Student Group, SYDE/MTE/NE)
2024 – 2026
Project: ASTRA Stratospheric Balloon Launch Team

Joshua Hill (Ugrad Co-op, URA, CS)
2024 – 2026
Project: Understanding In-Context Learning in Large Language Models

External Service

Peer review

Conference Area Chair: NeurIPS 2026 · ICML 2026 · FAccT 2026 · NeurIPS 2025 · ICML 2025 · NeurIPS 2024

Area Chair Recognition: NeurIPS 2025 Top Area Chair (top 10%)

Conference Program Committee (i.e. Reviewer): ACM EAAMO 2026 · ACM EAAMO 2025 · ICLR 2025 · AAAI 2025 · Machine Learning For Health Symposium (ML4H) 2024 · Conference on Health, Inference, and Learning (CHIL) 2024 · Canadian Artificial Intelligence Conference 2024 · ICML 2024 · NeurIPS 2023 · Conference on Health, Inference, and Learning (CHIL) 2023 · ACM FAccT 2023 · ICML 2023 · NeurIPS 2022 · ICML 2022 · ACM FAccT 2022 · ACM FAccT 2021 · NeurIPS 2021 · ICML 2021 · NeurIPS 2020 · ACM FAccT 2020 · ICML 2020 · NeurIPS 2019

Conference Ethics Reviewer: ICML 2026 · NeurIPS 2025 · ICML 2025 · NeurIPS 2024 · NeurIPS 2023 · NeurIPS 2022 · NeurIPS 2021

Workshop Proposal Reviewer: NeurIPS 2026

Workshop Reviewer: NeurIPS 2024 Workshop on Regulating Machine Learning · NeurIPS 2024 Workshop

on Algorithmic Fairness Through the Lens of Metrics and Evaluation • NeurIPS 2024 Workshop on Generative AI and Creativity • ICML 2024 Workshop on the Next Generation of AI Safety • AAAI 2024 Workshop on Recommendation Ecosystems: Modeling • Optimization and Incentive Design • NeurIPS 2023 Workshop on Robustness of Few-shot and Zero-shot Learning in Foundation Models • NeurIPS 2023 Workshop on Regulating Machine Learning • NeurIPS 2023 Workshop on Distribution Shifts: New Frontiers with Foundation Models • NeurIPS 2023 Workshop on Causal Representation Learning • ICML 2023 Workshop on Spurious Correlations, Invariance, and Stability • NeurIPS 2022 Workshop on Distribution Shifts • NeurIPS 2022 Workshop on Robustness in Sequence Modeling • NeurIPS 2022 Workshop on A Causal View on Dynamical Systems • NeurIPS 2022 Workshop on Algorithmic Fairness Through the Lens of Causality and Privacy • ICML 2022 Workshop on Continuous-time Methods for ML • ICML 2022 Workshop on Principles of Distribution Shifts • ICML 2022 Workshop on Spurious Correlations, Invariance, and Stability • NeurIPS 2021 Workshop on Distribution Shifts: Connecting Methods and Applications • NeurIPS 2021 Workshop on Algorithmic Fairness Through the Lens of Causality and Robustness • NeurIPS 2020 Workshop on Algorithmic Fairness Through the Lens of Causality and Interpretability • NeurIPS 2019 Fair Machine Learning for Health Workshop

Journal Action Editor: Transactions on Machine Learning Research (TMLR) • *2026 – Present*

Journal Reviewer: Journal of Machine Learning Research (JMLR)

Peer review of Funding Proposals Mitacs Accelerate Research Program

Meetings Organized

FAccT 2026 CRAFT Session: Common Ground (Type: CRAFT Session) June 25, 2026
Focus: closing the gap between algorithmic collective action research and LGBTQIA2S+ community needs (Role: Co-organizer) Montréal, QC, Canada

NeurIPS 2025 Workshop on Algorithmic Collective Action (Type: Workshop) December 6, 2025
Focus: AI user empowerment and coordination (Role: Lead Organizer) San Diego, CA, USA
Note: Accepted through NeurIPS peer review (19% acceptance rate); 50+ contributed paper submissions.

NeurIPS 2022 Tutorial on Algorithmic Fairness at the Intersections (Type: Tutorial) December 5, 2022
Focus: overlap of fairness, explainability, and robustness. (Role: Co-presenter) New Orleans, LA, USA

NeurIPS 2020 Workshop on Resistance AI (Type: Workshop) December 11, 2020
Focus: building human/AI systems that shift power to the people (Role: Co-organizer) Virtual

Internal Service

PhD Committees: Sebastian Parsons-Hall (UW ECE) • Majid Ghasemi (UW ECE) • Christian Schaible (UW ECE) • Soomin Shin (UW ECE) • Yiwei Lu (UW CS) • Sheng-Chieh Lin (UW CS) • Fangzheng Huang (UW ECE) • Kaixiang Zheng (UW ECE) • Zachary Birch (UW ECE) • Zening Li (UW ECE) • Da Saem Li (UW ECE) • Josh Sun (UW ECE)

Master's Committees: Guilherme Campos (UW ECE) • Zahra Khotanlou (UW ECE) • Kiernan McGuigan (UW SYDE) • Chang Liu (UW SYDE) • Harish Prakash (UW SYDE) • Arezoo Alipanah (UW ECE) • Xuanrui Zeng (UW ECE) • Thanushon Sivakaran (UW ECE) • Bavesh Balaji (UW SYDE) • Abdelrahman Elbadrawy (UW SYDE) • Adrian Chow (UW ECE) • Akash Karthikeyan (UW ECE)