Amir H. **Karimi**

ASSISTANT PROFESSOR | O'DONOVAN CHAIR IN TRUSTWORTHY AI | MACHING LEARNING SCIENTIST & ENGINEER

✉ amirhkarimi@gmail.com | ⌂ www.amirhkarimi.com | amirhkarimi_ | Google Scholar

*My career spans **top academic institutions** (Toronto, Waterloo, Stanford, ETH Zürich, Max Planck), **leading industry labs** (Google Brain, DeepMind, Meta AI), **major tech firms** (Meta, BlackBerry), and **startup ecosystems** (NEXT AI, The Next 36). This rare breadth gives me a deep appreciation for practical challenges, while my technical grounding and collaborations with world-class teams enable me to develop solutions that are both theoretically rigorous and real-world ready. My work focuses on explainable, trustworthy AI, and has been recognized with honors across research, industry, and teaching.*

Employment

University of Waterloo
Waterloo, CA

ASSISTANT PROFESSOR OF MACHINE LEARNING — *Sep 2023 - p.*

- Tenure-track Assistant Professor in the Electrical and Computer Engineering Department & School of Computer Science (cross-app.)
- PI of the **Collaborative Human-AI Reasoning Machines (CHARM) Lab**, dedicated to advancing safe and trustworthy human-AI teams
- **Affiliations:** Vector Institute Faculty Affiliate, Future of Life Institute Member of AI Safety Community Researchers

Google DeepMind
London, UK

RESEARCH SCIENTIST INTERN — *May 2022 - Oct 2022*

- Improved search via type- & IO-based neurally-guided program synthesis. Mentors: Lars Buesing, David Amos, Jessica Hamrick

Google Brain
Waterloo, CA

RESEARCH SCIENTIST INTERN — *Dec 2021 - Apr 2022*

- Investigated the causal effect of training hyperparameters on ML explanation. Mentors: Been Kim, Simon Kornblith
- Successfully published at ICML 2023, under the title *"On the Relationship Between Explanation and Prediction: A Causal View"* [C16]

Meta (Facebook) Inc.
New York, USA

SOFTWARE ENGINEER — *Aug 2015 - Sep 2016*

- Full-stack software engineer on the Enterprise Eng. team, responsible for front-end dev using React and hphp among others
- Implemented the landing page, a customizable notification framework, and the testing and logging platform for the Org Tool
- Successfully published at EMNLP 2016, under the title *"Key-Value Memory Networks for Directly Reading Documents"* [C2, P1]

Meta (Facebook) Inc.
Menlo Park, USA

SOFTWARE ENGINEER INTERN — *Feb 2014 - Apr 2014*

- Delivered a first-of-a-kind reporting tool for Facebook's largest Business Manager ad clients to view and manage historical ad budgets.
- Successfully reduced TTI < 7sec for Facebook's largest ad clients (100K+ ad accounts), by optimizing front-end JavaScript rendering.

BlackBerry Inc.
Toronto, CA

SOFTWARE ENGINEER INTERN — *May 2013 - Dec 2013*

- Implemented an automation framework to test 6,000+ ported selenium webdriver and WebKit tests for BB10.
- Successfully integrated browser team's test automation results with company central test database.

Stanford University
Stanford, USA

UNDERGRADUATE RESEARCH ASSISTANT - HIGH-FREQUENCY LAB — *May 2012 - Aug. 2012*

- Developed a helical antenna and used machine learning for precise 2D localization enabling hand gesture recognition.

Education

ETH Zürich & Max Planck Institute for Intelligent Systems
Zürich, CH & Tübingen, DE

PHD IN COMPUTER SCIENCE — *Oct 2018 - Jul 2023*

- **Thesis**: "Advances in Algorithmic Recourse: Ensuring Causal Consistency, Fairness, & Robustness"
- **Supervisors**: Prof. Bernhard Schölkopf & Prof. Isabel Valera
- **Major Awards**: NSERC CGS-D PhD Fellowship, Max Planck ETH PhD Fellowship, Google PhD Fellowship, ETH Zurich Medal

University of Waterloo
Waterloo, CA

MMATH IN COMPUTER SCIENCE — *Sep 2016 - Apr 2018*

- **Thesis**: "Exploring New Forms of Random Projections for Dimensionality Reduction"
- **Supervisors**: Prof. Alexander Wong & Prof. Ali Ghodsi
- **Major Award**: *Alumni Gold Medal* for highest standing across all master's programs at UWaterloo

University of Toronto
Toronto, CA

B.A.SC. IN ENGINEERING SCIENCE – ELECTRICAL AND COMPUTER STREAM — *Sep 2010 - Jun 2015*

- **Thesis**: "Benchmarking a Neuro-biologically Inspired Adaptive Controller"
- **Supervisors**: Prof. Chris Eliasmith & Prof. Richard Zemel
- **Major Award**: *Spirit of Engineering Science Award* for outstanding community contribution

Publications

My scholarly contributions on trustworthy, explainable, and causally-grounded AI have been ***showcased almost exclusively at top-tier AI and ML venues****—including NeurIPS, ICML, AAAI, AISTATS, ACM FAccT, and ACM AIES—where peer-reviewed conference proceedings are the primary venue for high-impact dissemination for artificial intelligence research. I have authored influential works such as a comprehensive survey in the prestigious ACM Computing Surveys, contributed a book chapter, and hold a U.S. patent. My research on algorithmic recourse has significantly shaped the field of responsible AI, helping elevate it from an emerging topic to a formal policy criterion—now mandated in Canada's Treasury Board Directive on Automated Decision-Making.* ***Several of my papers have each been cited over 100 times*** *and have been presented as spotlight and oral talks at top venues, reflecting sustained scholarly and policy impact. (📕 Google Scholar)*

Spotlight (💡) Oral (🎙) ≥ 100 Citations (☆) Best Paper (🏆) Patent (✹) Book Chapter (📕) Equal Contribution (*)

H-Index	G-Index	Max Citations	**Total Citations**
16	34	1,330 [C2]	**3,573**

|🎙|	|💡|	|☆|	|🏆|	|✹|	|📕|
11	3	7	1	1	1

Patents

P1	✹	2018	*US Patent* -	*"Key-Value Memory Networks"* Miller, Fisch, Dodge, **Karimi**, Bordes, Weston

Book Chapters

B1	📕	2022	*Springer LNAI* -	*"Towards Causal Algorithmic Recourse"* **Karimi**,* von Kügelgen,* Schölkopf, Valera

Journal Proceedings

J4	-	2026	*Trends in CogSci* I.F. 17.2	*"Imagining and building wise machines: The centrality of AI metacognition"* Johnson, **Karimi**, Bengio, Chater, Gerstenberg, Larson, Levine, Mitchell, Rahwan, Schölkopf, Grossmann
J3	-	2025	*Nature Scientific Reports* I.F. 3.9	*"Temporal Convolutional Transformer for EEG Based Motor Imagery Decoding"* Altaheri, Karray, **Karimi**
J2	☆	2022	*ACM Computing Surveys* I.F. 28.0	*"A survey of algorithmic recourse: contrastive explanations & consequential …* **Karimi**, Barthe, Schölkopf, Valera
J1	-	2011	*Optics Express* I.F. 3.2	*"Automated detection and density assessment of keratocytes in the human …* **Karimi**, Wong, Bizheva

Conference Proceedings

C18	-	2024	*ICML* A.R. %27.5	*"Prospector Heads: Generalized Feature Attribution for Large Models & Data"* Machiraju, Derry, Desai, Guha, **Karimi**, Zou, Altman, Ré, Mallick
C17	🎙	2023	*AAAI* A.R. %23.75	*"Causal Adversarial Perturbations for Individual Fairness and Robustness in …* Ehyaei, Mohammadi, **Karimi**, Samadi, Farnadi
C16	-	2023	*ICML* A.R. %27.9	*"On the Relationship Between Explanation and Prediction: A Causal View"* **Karimi**, Muandet, Kornblith, Schölkopf, Kim
C15	-	2023	*ICML* A.R. %27.9	*"On Data Manifolds Entailed by Structural Causal Models"* Dominguez-Olmedo, **Karimi**, Arvanitidis, Schölkopf
C14	-	2023	*FAccT* A.R. %24.6	*"Robustness Implies Fairness in Causal Algorithmic Recourse"* Ehyaei, **Karimi**, Schölkopf, Maghsudi
C13	☆ 💡	2022	*ICML* A.R. %21.9	*"On the Robustness of Causal Algorithmic Recourse"* Dominguez-Olmedo, **Karimi**, Schölkopf
C12	☆ 🎙	2022	*AAAI* A.R. %15.0	*"On the Fairness of Causal Algorithmic Recourse"* von Kügelgen, **Karimi**, Bhatt, Valera, Weller, Schölkopf
C11	🎙	2021	*ACM-AIES* A.R. %38.0	*"Scaling Guarantees for Nearest Counterfactual Explanations"* Mohammadi, **Karimi**, Barthe, Valera
C10	☆ 💡	2021	*ACM-FAccT* A.R. %25.0	*"Algorithmic Recourse: from Counterfactual Explanations to Interventions"* **Karimi**, Schölkopf, Valera
C9	☆ 💡	2020	*NeurIPS* A.R. %20.1	*"Algorithmic recourse under imperfect causal knowledge: a probabilistic …* **Karimi**,* von Kügelgen,* Schölkopf, Valera

C8	☆ 🎙	2019	*AISTATS* A.R. %32.4	*“Model-Agnostic Counterfactual Explanations for Consequential Decisions”* **Karimi**, Barthe, Balle, Valera
C7	🎙	2018	*IJCNN* A.R. %22.7	*“Distance Correlation Autoencoder”* Wang, **Karimi**, Ghodsi
C6	🎙	2018	*CVIS* A.R. ≤ %40.0	*“FEELS: a full-spectrum enhanced emotion learning system for assisting …* **Karimi**,* Boroomand,* Pfisterer, Wong
C5	🏆 🎙	2017	*CVIS* A.R. ≤ %40.0	*“Ensembles of Random Projections for Nonlinear Dimensionality Reduction”* **Karimi**, Shafiee, Ghodsi, Wong
C4	-	2017	*CCN* A.R. ≤ %40.0	*“Synthesizing Deep Neural Network Architectures using Biological Synaptic …* **Karimi**, Shafiee, Ghodsi, Wong
C3	🎙	2017	*ICIAR* A.R. ≤ %40.0	*“Discovery Radiomics via a Mixture Sequencers for Multi-Parametric MRI …* **Karimi**, Chung, Shafiee, Khalvati, Haider, Ghodsi, Wong
C2	☆ 🎙	2016	*EMNLP* A.R. %24.3	*“Key-Value Memory Networks for Directly Reading Documents”* Miller, Fisch, Dodge, **Karimi**, Bordes, Weston
C1	🎙	2016	*ICIP* A.R. ≤ %40.0	*“Spatio-temporal saliency detection using abstracted fully-connected …* **Karimi**, Shafiee, Scharfenberger, BenDaya, Haider, Talukdar, Clausi, Wong

Workshop Proceedings

W4	-	2025	*NeurIPS* -	*“Enhancing Algorithmic Recourse in Many-to-Many Multi-Agent Systems …* Khotanlou, **Karimi**
W3	🎙	2025	*EurIPS* -	*“Explainable AI is Causal Discovery in Disguise”* **Karimi**
W2	-	2018	*NeurIPS* -	*“Deep Variational Sufficient Dimensionality Reduction”* Banijamali, **Karimi**, Ghodsi
W1	-	2017	*NeurIPS* -	*“JADE: Joint Autoencoders for Dis-Entanglement”* **Karimi**,* Banijamali,* Ghodsi, Wong

Selected Pre-prints (as lead, senior, or core contributing author)

UP1	-	2025	- -	*“Bridging Brain with Foundation Models through Self-Supervised Learning”* Altaheri, Karray, Islam, Raju, **Karimi**

Advising

*One of my greatest privileges as an advisor is witnessing students grow into independent researchers. I have mentored and trained **23 highly-qualified personnel** (18 as faculty) in a diverse team—across culture, gender, and seniority—of Postdoctoral, PhD, Master’s, and undergraduate students, supporting their growth through interdisciplinary research and global collaboration. Several have gone on to top PhD programs, fellowships, and roles at leading institutions.*

Postdoc

2026-p.	**Supervisor**	Mohammad Hadi Sepanj (ECE)	-	🏛 Waterloo	-
2026-p.	**Co-supervisor**	Eugene Yu (PSYCH)	-	🏛 Waterloo	-
2024-5	**Co-supervisor**	Hamdi Altahery (ECE)	📘 J3, UP1	🏛 Waterloo	-

PhD

2026-p.	**supervisor**	Ahmed Abdelaal	-	🏛 Waterloo	-

Master’s

2024-p.	**Supervisor**	Zahra Khotanlou (ECE)	📘 W4	🏛 Waterloo	-
2025-p.	**Supervisor**	Hashir Ahmed (ECE)	-	🏛 Waterloo	-
2025-p.	**Supervisor**	Chenghao Tan (ECE)	-	🏛 Waterloo	-
2025	**Supervisor**	Dongzhuyuan Lu (ECE)	-	🏛 Waterloo	-
2025	**Supervisor**	Jieming Yu (ECE)	-	🏛 Waterloo	-
2024-5	**Co-supervisor**	Mina Kebriaee (ECE)	-	🏛 Waterloo	-
2025-p.	**Co-supervisor**	Hosna Oyarhoseini (CS) (**Vector Scholarship in AI**)	-	🏛 Waterloo	-
2024-p.	**Co-supervisor**	Maryam Ghorbansabagh (ECE)	-	🏛 Waterloo	-
2024	**Co-supervisor**	Zachary Wu (ECE)	-	🏛 Waterloo	-

Bachelors

2025-p.	**Supervisor**	Farzan Mirshekari (ECE)	-	🏛 Waterloo	-
2025-p.	**Supervisor**	Tom Wielemaker (MECH)	-	🏛 Waterloo	-
2025-p.	**Supervisor**	Hamza Mostafa (ECE)	-	🏛 Waterloo	→ Open AI Inc.
2024	**Supervisor**	Abubakar Bello (ECE)	-	🏛 Waterloo	→ Microsoft Inc.
2024	**Supervisor**	Mohammadreza Alavi (ECE)	-	🏛 Sharif	-

Mentoring (as a student)

2023-4	**PhD**	Ahmad Ehyaei	📖 C14	🏛 Tübingen	→ Intl. Max Planck Research Schools
2022-3	**PhD**	Miriam Rateike (**Google PhD Fellow 2023**)	-	🏛 Saarland	-
2021-3	**Master's**	Ricardo Dominguez-Olmedo (**Google PhD Fellow 2026**)	📖 C13, C15	🏛 Tübingen	→ Intl. Max Planck Research Schools
2019	**Master's**	Alexandra Walter	-	🏛 Tübingen	→ Helmholtz Data Sci. Sch. of Health
2020-2	**Bachelors**	Kiarash Mohammadi	📖 C11	🏛 Ferdowsi	→ MILA AI Institute

Honors, Funding, & Awards

*As an early-career researcher, I have been fortunate to receive strong support for my work—**securing over CAD $1,250,000** in competitive funding from generous sponsors including the University of Waterloo, Waterloo.AI, NSERC, Google, and CIFAR. I am honored to have **received several highly competitive national and international recognitions**, listed below.*

As a faculty

2025	**Sole-PI**	O'Donovan Chair in Trustworthy AI	(CAD $500,000)	*Waterloo, ON*
2025-6	**Lead-PI**	Mitacs Accelerate Entrepreneur Grant	(CAD $90,000) | 50%)	*Waterloo, ON*
2025-6	**Co-PI**	CIFAR Catalyst Grant	(CAD $100,000 | 33%)	*Waterloo, ON*
2024	**Co-PI**	Waterloo.AI Nexus of Data & AI Seed Funding	(CAD $20,000 | 50%)	*Waterloo, ON*
2024-8	**Sole-PI**	NSERC Discovery Grant (**highest amount awarded to an early-career researcher** in Waterloo Engineering in 2024)	(CAD $235,000)	*Waterloo, ON*
2024-8	**Sole-PI**	NSERC Discovery Grant Supplements	(CAD $12,500)	*Waterloo, ON*
2024	**Recipient**	**Igor Ivkovic Teaching Excellence Award** (1 of 1; competitive, student-nominated teaching award received in my first term	-	*Waterloo, ON*
2023	**Recipient**	Institution Startup Fund	(CAD $170,000)	*Waterloo, ON*

As a student

2023	**Recipient**	**ETH Zurich Medal** for Outstanding Doctoral Performance	(CHF 2,000)	*Zürich, CH*
2021	**Recipient**	**Google PhD Fellowship** in AI for Social Good (1 of 17 globally)	(USD $210,000)	*Tübingen, DE*
2018	**Recipient**	Max Planck ETH Center for Learning Systems PhD Fellowship	-	*Tübingen, DE*
2018	**Recipient**	NSERC Postgraduate Scholarship - Doctorate (PGS-D)	(CAD $63,000)	*Waterloo, CA*
2018	**Recipient**	**NSERC Canada Graduate Scholarship** - Doctorate (CGS-D)	(CAD $105,000) | Declined	*Waterloo, CA*
2018	**Recipient**	President's Graduate Scholarship (PGS)	(CAD $35,000) | Declined	*Waterloo, CA*
2018	**Recipient**	David R. Cheriton Graduate Scholarship	(CAD $20,000) | Declined	*Waterloo, CA*
2018	**Recipient**	**Alumni Gold Medal** – UWaterloo's top master's award (1 of 1)	-	*Waterloo, CA*
2015	**Recipient**	**Spirit of EngSci Award** for exemplary non-academic impact	-	*Toronto, CA*

Teaching

*Teaching is one of my greatest joys. I have had the privilege of educating diverse audiences—from (under)graduate students across Canada, Germany, and Switzerland to **over 40,000 learners online**—on topics such as machine learning and AI ethics. As a faculty member, I have taught **over 400 students in person** and was honored to receive the competitive, **student-nominated Igor Ivkovic Teaching Excellence Award in my first teaching term** at the University of Waterloo.*

2025	**Instructor**	Introduction to Machine Learning	∼ 140 students (grad.)	*Waterloo, CA*
2025	**Instructor**	Foundations of Computational Intelligence	∼ 100 students (ugrad.)	*Waterloo, CA*
2024	**Instructor**	Introduction to Machine Learning	∼ 10 students (intl.)	*online*
2024	**Instructor**	Tools of Intelligent Systems Design	∼ 140 students (grad.)	*Waterloo, CA*
2024	**Instructor**	Foundations of Computational Intelligence (**Igor Ivkovic Teaching Excellence Award**)	∼ 90 students (ugrad.)	*Waterloo, CA*
2022-p.	**Co-instructor**	Providing free public education on basic & advanced AI subjects on YouTube, Instagram, Substack, & Medium	**≥ 40,000 students**	*online*

Invited Talks (Selected)

“Explainable AI is Causality in Disguise” EurIPS Theory of Explainable AI (2025)

“Building Bridges: Towards Trustworthy Human-AI Decision Making” Vector Institute (2024), Toronto Machine Learning Summit (2024), Waterloo Alumni Reunions (2024, 2025), Waterloo.AI Seminars (2024), Waterloo.AI Nexus of Data & AI Event (2024), Waterloo VIP Lab (2024), Waterloo ECE Seminars (2024)

“Algorithmic recourse: from theory to practice” Google Brain LUMA team (2022), Google DeepMind (2022), MILA (2022), IMS Annual Meeting (2022), Harvard University (2021), NEC Europe Labs (2021), Cyber Valley Health (2021), ETH IML Seminars (2020), UCL Causality Group (2020)

Panelist:

2024	**Tech Horizons Executive Forum Panel on Trust in AI**	∼ 300 attendees	*Toronto, CA*
2024	BBC Radio 4 Episode on AI for Emotion Detection	-	*online*
2023	AI and the Transformation of Social Science Research	∼ 50 attendees	*Waterloo, CA*

Scientific Reviewing

Reviewer (Journals & Grants)

2024-p.	Journal of Artificial Intelligence, Journal of Machine Learning Research, NSERC Mitacs Accelerate, NSERC Discovery Grants

Program Committee (Conference Reviewer)

2026	ICML
2025	ICLR, NeurIPS, AISTATS
2023	ICLR
2022	ICML, AISTATS
2021	ICLR, NeurIPS
2020	ICLR, NeurIPS, ACM FAccT, ECML, AISTATS

Other Service

Workshop & Symposium Organization

2023	**Co-organizer**	ICML Workshop on Counterfactuals in Minds & Machines	∼ 50 attendees	*Hawaii, USA*
2021	**Co-organizer**	ELLIS Workshop on Causethical ML	∼ 50 attendees	*online*
2021	**Co-organizer**	ICML Workshop on Algorithmic Recourse	∼ 50 attendees	*online*
2020	**Co-organizer**	NeurIPS Symposium on Algorithmic Recourse	∼ 100 attendees	*online*
2020	**Support**	Machine Learning Summer School (MLSS)	∼ 150 attendees	*online*

Program & Fellowship Selection

2024-p.	**Reviewer**	ICML Workshops Proposal Review Committee	//	*online*
2024-p.	**Reviewer**	Vector Scholarship in AI Review Committee	//	*Toronto, CA*

Departmental & Faculty Service

2024-p.	**Dept. Repr.**	Engineering Faculty Council	-	*Waterloo, CA*

Graduate Examination & Committee Service

2024-p.	**Chair**	PhD Comprehensive Exam	//	*Waterloo, CA*
2024-p.	**Examiner**	PhD Defence	/	*Waterloo, CA*
2024-p.	**Examiner**	PhD Comprehensive Seminar	/	*Waterloo, CA*
2024-p.	**Examiner**	PhD Comprehensive Proposal Exam	//	*Waterloo, CA*
2024-p.	**Examiner**	PhD Comprehensive Background Exam	𝍸 //	*Waterloo, CA*
2024-p.	**Examiner**	Master’s Seminar	//	*Waterloo, CA*

References

Professor & Director	**Bernhard Schölkopf**	MPI for Intelligent Systems	sekretariat-schoelkopf@tue.mpg.de
Professor	**Isabel Valera**	Department of CS, Saarland University	ivalera@cs.uni-saarland.de
Professor & Director	**Gilles Barthe**	MPI for Security and Privacy	gilles.barthe@mpi-sp.org
Senior Staff Research Scientist	**Been Kim**	Google DeepMind	beenkim@google.com
Assistant Professor	**Himabindu Lakkaraju**	Department of CS, Harvard University	hlakkaraju@seas.harvard.edu