School of Philosophy
Centre for Technomoral Futures
University of Edinburgh
atoosa.kasirzadeh@ed.ac.uk
http://kasirzadeh.org

Atoosa Kasirzadeh

Last updated June 2023

Employment

Jan 2022– **Assistant Professor**, University of Edinburgh, UK
- *Chancellor's Fellow*
- *Director of Research* at the Centre for Technomoral Futures, Edinburgh Futures Institute

Aug–Dec 2021 **Visiting Research Scientist**, DeepMind, UK

2019–2021 **Postdoctoral Fellow**, Australian National University (Humanizing Machine Intelligence)

Non-academic Appointments

Jan 2023– **Research Lead**, The Alan Turing Institute, UK

Jan 2023 – **Senior Policy Fellow**, DCMS/UKRI

Education

2021 **PhD, Philosophy of Science & Technology**, University of Toronto, Canada

2015 **PhD, Mathematics**, Ecole Polytechnique of Montreal, Canada
- *Researcher in the Group for Research in Decision Analysis (GERAD)*

2009 **M.Sc in Systems Engineering**, Royal Institute of Technology, Stockholm, Sweden

Visiting Positions

2019, 22, 23 **Visiting scholar**, Munich Center for Mathematical Philosophy, Germany

2017 **Visiting scholar**, University of Paris 7-Diderot, France

2009 **Visiting scholar**, Transport & Mobility Laboratory, EPFL, Switzerland

Refereed Publications

Published/Accepted

2023 "User Tampering in Reinforcement Learning Recommender Systems" (with Charles Evans), *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES)*

2023 "Typology of Risks of Generative Text-to-Image Models" (with Eddie Ungless and Charlotte Bird), *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES)*

2023 "Reconciling Governmental Use of Online Targeting With Democracy" (with Katja Andrić), *ACM Proceedings of Fairness, Accountability, and Transparency (FAccT)*

2023 "Science in the Age of Large Language Models" (with Abeba Birhane, David Leslie, Sandra Wachter), *Nature Reviews Physics*

2023 In Conversation with Artificial Intelligence: Aligning Language Models with Human Values (with Iason Gabriel), *Philosophy and Technology*

2022 "Taxonomy of Risks Posed by Language Models" (with Laura Weidinger et al.), *ACM Proceedings of Fairness, Accountability, and Transparency (FAccT)*

2022 "Algorithmic Fairness and Structural Injustice: Insights from Feminist Political Philosophy", *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES)*

2022 "Algorithmic and Human Decision-Making: For a Double Standard of Transparency" (with Mario Günther), *AI & Society*

2022 "Counter Countermathematical Explanations", *Erkenntnis*

2021 "A New Role for Mathematics in Empirical Sciences", *Philosophy of Science*

2021 "The Ethical Gravity Thesis: Marrian Levels and the Persistence of Algorithmic Bias in Automated Decision-making Systems" (with Colin Klein), *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES)*

2021 "Fairness and Data Protection Impact Assessments" (with Damian Clifford), *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES)*

2021 "Reasons, Values, and Stakeholders: A Philosophical Framework For Explainable Artificial Intelligence", *ACM Proceedings of Fairness, Accountability, and Transparency (FAccT)*

2021 "The Use and Misuse of Counterfactuals in Ethical Machine Learning", (with Andrew Smart), *ACM Proceedings of Fairness, Accountability, and Transparency Conference (FAccT)*

2017 "Airline Crew Scheduling: Models, Algorithms, Data Sets", A. Kasirzadeh, M. Saddoune M, F. Soumis, *EURO Journal on Transportation and Logistics, p. 1–27*

Invited Articles

2024 "Manipulative Explanations in AI" (with Emily Sullivan), Book Chapter in *Philosophy of Science for Machine Learning: Core Issues and New Perspectives, Synthese Library.*

2024 "Intelligent Capacities in Artificial Systems" (with Victoria McGeer), Book Chapter in *Artificial Dispositions: Investigating Ethical and Metaphysical Issues, Bloomsbury Publishing.*

Reviews

2019 "Applying Mathematics: Immersion, Inference, Interpretation, Otavio Bueno and Steven French", (Book Review with James Robert Brown) in *Philosophy of Science*

2018 "Scientific Collaboration and Collective Knowledge, Thomas Boyer-Kassem et al.", (Book Review) in *British Journal for the Philosophy of Science Review of Books*

Selected Honors, Awards, Grants

2023 AHRC/DCMS Senior Policy Fellowship (£60,000)

2022 Alan Turing Institute Grant for "New Perspective on AI Futures" workshop (PI, £25,000)

2022 AHRC/SGSAH Collaborative Doctoral Award (Co-I, £63,624)

2019 ANU Early Career Research Travel Grant

2018 Ethics of AI Graduate Research Fellowship

2018 Philosophy of Science Association (PSA) Travel Grant

2016–2019 Faculty of Art & Science Top Doctoral Fellowship (FAST), University of Toronto

2016, 2018 Full Stipend International Rationality Summer Institute

2015–2021 University of Toronto Doctoral Fellowship

2014 Canadian Mathematical Society for Women in Mathematics Travel Award

2009–2015 Ecole Polytechnique de Montreal Doctoral Fellowship

Advising

PhD Students

2022– Charlotte Bird (University of Edinburgh; co-advised with Eva Luger)

2022– Matthew Wragg (University of Edinburgh; co-advised with Nick Teanor)

2021– Alexander Martin Mussgnug (University of Edinburgh; co-advised with Shannon Vallor)

MSc Students

2023 Hannah Mayo (University of Edinburgh)

2022 Katja Andrić (University of Edinburgh)

2022 Alice Gibson-sey (University of Edinburgh)

2022 Sabrina Pate (University of Edinburgh)

Undergraduate Students

2020, 2021 Charles Evans (Australian National University)

Service to the Profession

2022–2023 Program chair for 2023 ACM Conference on Fairness, Accountability, and Transparency

2022–2023 Member of the Cross-College Research Ethics Advisory Group for the Global Data Institute for Child Safety

2022 Invited Lecturer at Summer School for Women in Philosophy, Munich, Germany

Since 2015 Reviewer for MIT Press, Oxford University Press, Swiss National Science Foundation, Nature Machine Intelligence, Mind, Philosophical Studies, Philosophy of Science, British Journal for the Philosophy of Science, Philosophy and Technology, The Journal of Political Philosophy, Erkenntnis, Minds and Machines, Studies in History and Philosophy of Science, NeurIPS 2020 (Ethics Section), Organon F, Annals of Operations Research, Optimization Letters

2022 Program Committee Member for the British Society for the Philosophy of Science

2021, 22 Program Committee Member for ACM Conference on Fairness, Accountability, and Transparency

2021, 22, 23 Program Committee Member for the AAAI/ACM Conference on AI, Ethics, and Society

2021–2022 Leading Co-editor for a Special Issue of *Synthese* on Philosophy of Science in light of Artificial Intelligence

2020, 21, 23 Ethical Reviewer for NeurIPS (Top-tier Publication Venue in Machine Learning)

2019 Co-organizer of Symposium on Non-causal Explanations at Canadian Society for the History and Philosophy of Science with Nicholas Fillion

2018–2019 Executive Member of Canadian Society for Women in Philosophy (CSWIP)

Selected Public Outreach

2023 Invited Panelist for "Beyond the Code: Think for Tomorrow", Sziget Festival, Hungary

2023 "Can regulating AI suppress innovation?", Aljazeera English Inside Story

2023 "Governing and regulating generative AI models", Aljazeera English Inside Story (alongside Gary Marcus and Ramesh Srinivasan)

2023 Invited Contribution to DailyNous "ChatGPT, Large Language Technologies, and the Bumpy Road of Benefiting Humanity"

2023 Invited Contribution to the Royal Society of Edinburgh Magazine on Ethics of AI "The socio-ethical challenges of generative artificial intelligence"

2022 Blogpost for Google DeepMind on the Alignment of Large Language Models

2022 Interview with the *ABC Radio* about AI Ethics and the Future

2022 Interview with *the Times* about the Sentience of Large Language Models

2020 Invited Panelist for "Taming the Terminator: Law, Ethics, and Artificial Intelligence"

Teaching

Instructor

2022 Ethics and Politics of Data, University of Edinburgh (with Shannon Vallor)

2022 Explanation in Computational and Mathematical Sciences, LMU Munich (Summer School)

2020, 2021 Ethical and Societal Implications of Artificial Intelligence, Australian National University

2019 Philosophy of Technology, University of Toronto (Proposal Winner for a New Course)

2018 Ethics of Artificial intelligence, LMU Munich and Venice International University (With Dr. Fiorella Battaglia)

Guest Lecturer

2022 Democracy and Artificial Intelligence; University of Cambridge, UK

2022 Ethics and Large Language Models; University of Cambridge, UK

2022 Rationality and Explainable Artificial Intelligence, International Rationality Summer Institute, Landau, Germany (Summer School)

2020 Bias and Fairness in Machine Learning, Decision, Fairness, and Machine Learning, University of Bayreuth, Germany

2020 Bias and Fairness in Machine Learning, Advanced Computing Research and Development Methods, Australian National University

2020 Introduction to History and Philosophy of Science, Advanced Computing Research and Development Methods, Australian National University

2018 Moral Status of Artificial intelligence, Minds and Machines, University of Toronto

University of Toronto TA Assignments

2021 Introduction to Bioethics

2020 Social Implications of Information Technology

2020 Introduction to History & Philosophy of science

2018 Introduction to Moral Philosophy: Persons and Values

2018 History & Philosophy of Evolutionary Biology

2017 History of Physics

2017 Probability & Inductive Logic

2016, 2017 Minds & Machines (Philosophy of Artificial Intelligence)

2016 Making Sense of Uncertainty

Academic Talks

"In conversation with Artificial Intelligence: aligning language models with human values"

Mar 2023 - *Carnegie Mellon University, Pittsburgh, US*

Jan 2023 - *Philosophy, AI, and Society Workshop, Stanford University, US*

Nov 2022 - *Centre for Advancing Responsible & Ethical Artificial Intelligence, University of Guelph, Canada*
Oct 2022 - *Ethics of AI workshop, University of Notre Dame Rome Global Gateway, Italy*
June 2022 - *Workshop on AI in Science, Cambridge-LMU, Germany*

Invited Presidential panelist on "Automating Science"
Nov 2022 - *Philosophy of Science Association (PSA), Pittsburgh, USA*

"Style of Reasoning of Machine Learning"
Nov 2022 - *Philosophy of Science Association (PSA), Pittsburgh, USA*

"Algorithmic Fairness in Light of Structural Injustice"
Apr 2022 - *Philosophy Colloquium Series, Cornell University, USA*

"Responsible Algorithmic Fairness: Insights from Feminist Political Philosophy"
May 2022 - *Workshop on Responsibility ans AI, University of Vienna, Austria*
Mar 2022 - *Scuola Superiore Sant'Anna di Pisa, Italy*

"Kinds of Explanation in Machine Learning"
Nov 2021 - 27^{th} *Biennial Meeting of the Philosophy of Science Association, Baltimore, USA*
Nov 2021 - *Philosophy of Science Meets Machine Learning Workshop, Tübingen, Germany*

"The Ethical Gravity Thesis: Marrian Levels and the Persistence of Algorithmic Bias in Automated Decision-making Systems"
May 2021 - *AAAI/ACM Conference on AI, Ethics, and Society*
May 2021 - *Virtual Philosophy of Data Science Symposium, Upstate Workshop on AI and Human Values*

"Reasons, Values, and Stakeholders: A Philosophical Framework For Explainable Artificial Intelligence"
Mar 2021 - *ACM Conference on Fairness, Accountability, and Transparency*

"Algorithmic and Human Decision-making: Arguments for a Double Standard of Transparency"
March 2021 - *International Ethics of Data Science Conference, Sydney, Australia*

"Counter Countermathematical Explanations"
May 2022 - *Conference on Mathematical Explanation and Understanding, Paris, France*
Jan 2022 - *Vrije Universiteit Brussel, Belgium*
Jan 2021 - *Eastern APA Meeting, New York, USA*

"The Use and Misuse of Counterfactuals in Fair Machine Learning"
Nov 2021 - 27^{th} *Biennial Meeting of the Philosophy of Science Association, Baltimore, USA*
Mar 2021 - *ACM Conference on Fairness, Accountability, and Transparency*

Dec 2020 - *Algorithmic Fairness through the Lens of Causality and Interpretability, NeurIPS Workshop, Vancouver, Canada*

Oct 2020 - *Virtual Workshop on the Philosophy of Medical AI, Tübingen, Germany*

Sep 2020 - *Bias and Fairness in AI Workshop at ECMLPKDD 2020, Ghent, Belgium*

"Why and How Explanations Matter: Connecting Moral and Technical Dimensions of Algorithmic Decision-making"

July 2020 - *Interpretable Machine Learning, Simons Institute, University of California Berkeley, USA*

June 2020 - *Athena in Action Virtual Workshop, Department of Philosophy, Cornell University, USA*

''Mathematical and causal faces of explainable Artificial Intelligence"

Dec 2019 - *NeurIPS Workshop on Human-centric Machine Learning, Vancouver, Canada*

Nov 2019 - *University of Washington Department of Philosophy, Seattle, US*

Nov 2019 - *University of California San Diego Department of Philosophy, US*

Nov 2019 - *Irvine Department of Logic and Philosophy of science, US*

Oct 2019 - *Stanford Department of philosophy, US*

Oct 2019 - *Australian National University, Department of philosophy*

Oct 2019 - *Australian National University, Department of Computer Science*

''Levels and a new role for mathematics in empirical sciences"

July 2019 - *British Society for the Philosophy of Science, Durham, UK*

June 2019 - *Canadian Society for the History and Philosophy of Science, Vancouver, Canada*

''Can Mathematics Really Make a Difference?"

May 2019 - *Non-Causal Explanations: Logical, Linguistic and Philosophical Perspectives, Ghent, Belgium*

Jan 2019 - *Workshop on Metaphysical Explanation in Science, Birmingham, UK*

Feb 2019 ''Ethics, Explanations, Machine Learning" (Poster)
- *Canada-United Kingdom Symposium On Ethics In Artificial Intelligence, Ottawa, Canada*

"Computer Simulations & the Production of New Pragmatic Knowledge in Social Sciences''

Nov 2018 - 26^{th} *Biennial Meeting of the Philosophy of Science Association, Seattle, USA*

Aug 2017 "Auxiliary Probabilities & the Principal Principle''
- 9^{th} *European Congress of Analytic Philosophy, Munich, Germany*

Jun 2017 "Non-causal & Mathematics-based Explanations''
- *Workshop on Causation, Explanation, and Conditionals, Tutzing & Munich, Germany*

May 2017 "Computer Simulations, Mathematical Models, and Production of New Knowledge''
- *Canadian Society for the History and Philosophy of Science (CSHPS), Toronto, Canada*

Apr 2016 "The Explanatory Role of Mathematics in Social Sciences: the Case of Health Economics''

- 43^{rd} *Annual Philosophy of Science Conference, Dubrovnik, Croatia (invited)*

"An Integrated Simultaneous Approach for Pilots and Copilots Rescheduling Problem"

Aug 2015 - *Multidisciplinary International Scheduling Conference (MISTA), Prague, Czech Republic*

June 2015 - *Joint International Meeting of Canadian Operational Research Society (CORS) & Institute for Operations Research and the Management Sciences (INFORMS), Montreal, Canada (invited)*

"Simultaneous Optimization of Personalized Integrated Scheduling for Pilots and Copilots"

Nov 2014 - *Institute for Operations Research and the Management Sciences (INFORMS) Annual Meeting, San Francisco, USA (invited)*

Oct 2014 - *Canadian Mathematical Society Women in Math, Banff International Research Center for Mathematical Innovation and Discovery, Banff, Canada*

July 2014 - 20^{th} *Conference of the International Federation of Operational Research Societies (IFORS), Barcelona, Spain*

May 2014 - *Optimization Days, Montreal, Canada*

"Integrated Personalized Crew Pairing and Crew Assignment Problem"

July 2013 - 26^{th} *European Conference on Operational Research, Rome, Italy*

"Integrated Personalized Crew Pairing and Crew Assignment Problem"

May 2013 - *Optimization Days, Montreal, Canada*

Commentaries

2018 "There sweep great general principles which all the laws seem to follow" (Marc Lange)
- *HPS Colloquium series, University of Toronto, Canada*