

2024

FRONTIERS
OF

SCIENCE

ARTIFICIAL INTELLIGENCE

THE ROYAL SOCIETY OF CANADA

AND

THE ROYAL SOCIETY

OTTAWA, ONTARIO, CANADA

FEB. 13-15



THE
ROYAL
SOCIETY

THANK YOU TO OUR PARTNERS



National Research
Council Canada

Conseil national de
recherches Canada

Dear Colleagues,

We are very pleased to welcome you to the 2024 UK-Canada Frontiers of Science Meeting on Artificial Intelligence. This auspicious event will bring together current and future leaders in AI from the UK and Canada for collaborative discussions around the coming *frontier* of this important topic. The event will mark the second Frontiers of Science meeting held between the Royal Society and the Royal Society of Canada, which was previously organized in Whistler in 2016.

Artificial Intelligence is a hot topic of conversation, even more so than when this event was first conceived. A multi-faceted issue, important conversations are being had (and not had) about AI's place in our lives, with questions of regulation, policy, ethics, and ensuring AI for the common good. To explore these ideas we have organized this week's event around four themes: *AI and ethics, AI and Arts, AI and Health, and AI and Deep Learning*. By building our event around these core themes we will ensure that we are having relevant conversations about the coming frontier.

We are so grateful that you are able to join us this week here in Ottawa. We are looking forward to an exciting event.

Sincerely,

Ian Wiggins, Director of International Affairs, The Royal Society
Paul Young, International Secretary, The Royal Society of Canada



Ian Wiggins
Director of International Affairs
The Royal Society



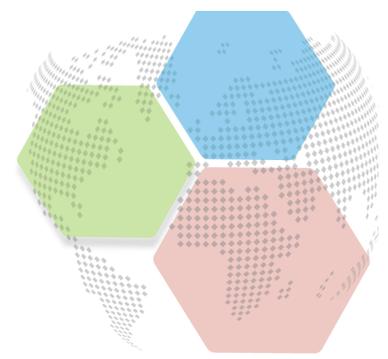
Paul Young
International Secretary,
Royal Society of Canada

#UKCanFoS24



FRONTIERS OF SCIENCE: ARTIFICIAL INTELLIGENCE

THE ROYAL SOCIETY OF CANADA
AND THE ROYAL SOCIETY
OTTAWA, ONTARIO, CANADA



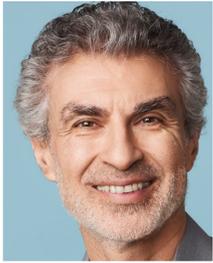
SCHEDULE

MONDAY, FEBRUARY 12	
6:00 pm	Evening Reception Fairmont Chateau Laurier, Zoé's
TUESDAY, FEBRUARY 13	
8:00 am	Breakfast Begins Fairmont Chateau Laurier, Québec Room
9:00 am	Welcoming Remarks
	Territorial Acknowledgment with Grandma Karen MacInnis , Knowledge Keeper and Elder
9:30 am	KEYNOTE PRESENTATION
	Yoshua Bengio , Department of Computer Science and Operations Research, Université de Montreal
	Why and How Could we Design Aligned and Provably Safe AI? This presentation will start by presenting the major risks associated with advanced AI systems, as we progress towards human-level competence - or more - on more and more cognitive tasks. This suggests two significant challenges to make sure that we avoid catastrophic outcomes: the technical alignment challenge (finding a way to design AI that behaves as intended, a yet unsolved problem) and the political challenge of coordination and democratic governance, at national and international levels (making sure that everyone follows the safety protocols and does not abuse the power of AI). I will then outline a machine learning research program to obtain quantitative and conservative risk evaluations to address the alignment challenge. This is based on deep learning methods for amortized inference of the Bayesian posterior over causal theories of the data available to the AI, and taking a risk management posture in which the AI takes conservative decisions accounting for the uncertainty around consequences of its actions and what humans really consider harmful.
10:15 am	Group Discussion
<i>Coffee served</i>	
11:00 am	The Great Debate Hosted by Yoshua Bengio 1. Do we have the necessary mechanisms for participatory action research on AI? 2. What is the scope of education needed to support AI literacy? And do we have a sense that there is a critical mass of the scientific community that is rising up to this challenge? 3. How can we promote an academic research environment that can tackle today's AI scale? Can we really justify continued development of large models?
12:00 pm	Lunch
1:00 pm	Breakout Groups The Impact of AI: AI and Ethics, AI and Health, AI and Privacy, AI and Deep Learning Problems include privacy concerns and the effect of energy consumption, among others. On the other hand, AI could lead to advancements in areas such as drug discovery and osmological analysis. We need to consider how both society and academia can develop AI while ensuring overall societal benefits. A sub-topic within this context is that even within academia and science, sometimes techniques are not discussed in a transparent and honest manner (e.g. glorying performance metrics) which makes them less reliable when used in practice, but also makes it difficult to apply them in science. What can we do as AI and data scientists to ensure transparency and accessibility? How do we foster easier comparison to enable access to AI technology within science? Perhaps a starting point is identifying cases where there is more benefit in using simpler models than directly jumping to deep learning.
2:30 pm	Breakout Group Presentations
4:00 pm	Cultural Activity (Skating, Ice Sculptures, Winterlude, Beavertails)
6:30 pm	Dinner Métropolitain Brasserie



WEDNESDAY, FEBRUARY 14		
8:00 am	Breakfast Fairmont Chateau Laurier, Québec Room	
9:00 am	PANEL PRESENTATIONS AI and its impacts, including regulatory aspects ?	
	Moderator Elissa Strome , Executive Director, Pan-Canadian AI Strategy, CIFAR	
	Panelists	Stephanie Dick , Simon Fraser University, Canada
		Atoosa Kasirzadeh , University of Edinburgh and Alan Turing Institute, UK
		Luke Stark , Western University, Canada
Erin Young , The Alan Turing Institute, UK		
10:00 am	Group Discussion	
11:00 am	KEYNOTE PRESENTATION	
	Doina Precup , Machine decision-making, McGill University	
	Learning how to take decisions from preferences Learning how to take sequences of decisions under uncertainty is a hallmark of intelligence. Reinforcement learning, which draws inspiration from neuroscience, animal learning theory, control theory and economics, has been the method of choice for learning how to predict and act from interaction with an unknown environment. In reinforcement learning, the goal of the agent is expressed as the maximization of an expected numerical reward signal. However, when AI agents are trained by interacting with people, eliciting a numerical reward can be difficult and noisy. As a result, the method of choice is to allow a person to provide a preference between two different trajectories (for example, two answers to a query). But is this approach really justified? In this talk, I will discuss recent work in which we show that preferences can indeed be more expressive than numerical rewards, while still allowing for the existence of optimal behaviors. I will discuss recent algorithmic advances for learning such behaviors from data, in reinforcement-learning-style. Finally, I will discuss why alignment should be viewed as a sequential decision-making problem, rather than a one-step process, and why I am optimistic about the future AI-human coexistence.	
12:00 pm	Lunch	
1:00 pm	Breakout Groups Collaboration Between Countries and Disciplines: AI and Ethics; AI and Health; AI and Privacy; AI and Deep Learning	
	Interdisciplinary work is often difficult to publish (and fund) because it requires reviewing expertise from multiple distinct communities. How can we encourage collaborations across disciplines while supporting the careers of researchers involved in these collaborations (e.g., students, early-career faculty)? Different countries have distinct funding and recruiting priorities. How can we decrease the overhead of collaboration across countries to facilitate research projects that leverage each country's strengths? Are the benefits of interdisciplinary collaboration immediately clear to your areas? For example, in biology, researchers may benefit a lot by having AI collaborators, and they might also have possible venues for publication, whereas mainstream AI researchers will typically go for mainstream AI conferences, where an application in other areas <u>has to</u> be written in a way that's readable and accessible to a computer science audience. What other ways have people worked this out?	
2:30 pm	Breakout Group Presentations	
4:00 pm	Policy Discussion David Prodger, Denisse Albornoz and Nicole Mwananshiku	
5:00 pm	Light Dinner	
6:30pm	High Commission Reception Ottawa Art Gallery (OAG) – Canopés and Drinks reception	
THURSDAY, FEBRUARY 15		
8:00 am	Breakfast Fairmont Chateau Laurier, Québec Room	
9:00 am	Morning Wrap-up Co-Chairs and Panelists	
10:00 am	Reflection Discussion	
11:00 am	Closing Ceremony	
12:00 pm	Travel Out	

KEYNOTE PRESENTATIONS



Yoshua Bengio (He/Him)
Université de Montreal, Canada

Yoshua Bengio is Full Professor in the Department of Computer Science and Operations Research at Université de Montreal, as well as the Founder and Scientific Director of Mila and the Scientific Director of IVADO. He also holds a Canada CIFAR AI Chair. Considered one of the world's leaders in artificial intelligence and deep learning, he is the recipient of the 2018 A.M. Turing Award with Geoff Hinton and Yann LeCun, known as the Nobel prize of computing.

Fellow of both the Royal Society of London and Canada an Officer of the Order of Canada and member of the UN's Scientific Advisory Board for Independent Advice on Breakthroughs in Science and Technology. In 2023 Yoshua Bengio received the Gerhard Herzberg Canada Gold Medal for Science and Engineering.



Doina Precup (She/Her)
McGill University, DeepMind, CIFAR AI Chair, Canada

Doina Precup splits her time between McGill University / Mila, where she holds a Canada-CIFAR AI chair, and Google DeepMind, where she is a Research Director. Her research interests are in the areas of reinforcement learning, reasoning under uncertainty, time series analysis, and diverse applications of machine learning in areas that have social impact, such as health care. In 2022, she was elected Fellow of the Royal Society of Canada. She is also a senior fellow of CIFAR's Learning in Machines and Brains program. Doina Precup has played a leadership role

in the Canadian AI ecosystem, helping to establish Mila and chairing the National Program Committee of the CIFAR Pan-Canadian AI Strategy until 2022. She also co-founded the AI4Good lab, an initiative to increase gender diversity in the AI and machine learning workforce.

BIOGRAPHIES



Co-Chair, Vaishak Belle (He/Him)
University of Edinburgh, UK

Dr Vaishak Belle is Reader at the University of Edinburgh, an Alan Turing Fellow, and a Royal Society University Research Fellow. He has made a career out of doing research on the science and technology of AI. He has published close to 100 peer-reviewed articles, won best paper awards, and consulted with banks on explainability. As PI and Col, he has secured a grant income of close to 8 million pounds.



Co-Chair, Nicolas Papernot (He/Him)
University of Toronto & RSC College, Canada

Nicolas Papernot is an Assistant Professor of Computer Engineering and Computer Science at the University of Toronto. He also holds a Canada CIFAR AI Chair at the Vector Institute, and is a faculty affiliate at the Schwartz Reisman Institute. His research interests span the security and privacy of machine learning. Some of his group's recent projects include generative model collapse, cryptographic auditing of ML, private learning, proof-of-learning, and machine unlearning. Nicolas is an Alfred P. Sloan Research Fellow in Computer Science and a Member

of the Royal Society of Canada's College of New Scholars. His work on differentially private machine learning was awarded an outstanding paper at ICLR 2022 and a best paper at ICLR 2017. He co-created the IEEE Conference on Secure and Trustworthy Machine Learning (SaTML) and is co-chairing its first two editions in 2023 and 2024. He previously served as an associate chair of the IEEE Symposium on Security and Privacy (Oakland), and an area chair of NeurIPS. Nicolas earned his Ph.D. at the Pennsylvania State University, working with Prof. Patrick McDaniel and supported by a Google PhD Fellowship. Upon graduating, he spent a year at Google Brain where he still spends some of his time.





Alfie Abdul-Rahman (She/Her)
King's College London, UK

Alfie Abdul-Rahman is a Senior Lecturer in Computer Science at King's College London. She received her PhD from Swansea University in Computer Science. Before joining King's College London, she was a Research Associate at the University of Oxford e-Research Centre. She worked as a Research Engineer in HP Labs Bristol on document engineering, and then as a Software Developer in London, working on multi-format publishing. Her research interests include data and information visualization, computer graphics, human-computer interaction, and digital humanities.



Denisse Albornoz (She/Her)
The Royal Society, UK

Denisse Albornoz is Senior Policy Adviser in the Data and Digital Technologies team (Science Policy section) at the Royal Society. The team focuses on how artificial intelligence and other data-driven technologies can, and should, be used to benefit humanity. She oversees the following projects: Science in the Age of AI, (on the impact of AI on scientific research and policy implications) Data for Emergencies (on the use of open, trusted, and resilient data systems in times of emergencies) and the Online Information Environment (how data-driven technologies are changing the nature of scientific misinformation).



Luca Arnaboldi (He/Him)
University of Birmingham, UK

Luca Arnaboldi is an Assistant Professor at the University of Birmingham, within the Centre for Security and Privacy. His research focus is cybersecurity, machine learning and formal verification. He has published several papers, both in the domain of machine learning solutions for security and security of machine learning systems. He was awarded his PhD at Newcastle University on the topic of improving explainability and adaptability of attack detection systems. His most recent work is on the topic of formal verification of security in machine learning systems; where he has contributed with partner institutions on the topic of autonomous vehicle security, formal verification of neural networks, AI ethics, security compliance, and natural language processing.



Colin Bellinger (He/Him)
National Research Council, Canada

Dr. Colin Bellinger is a Research Officer at the National Research Council of Canada and an Adjunct Professor at Dalhousie University in the Faculty of Computer Science. His research leverages real-world problems in science, health, security and industry as a gateway to understand how machine learning algorithms learn from and are impacted by adverse and limited data domains. Dr. Bellinger's current research focuses on developing robust, data efficient supervised learning and reinforcement learning algorithms with applications in automated chemistry and materials design, where samples are limited and costly.



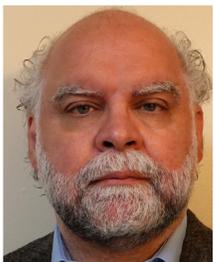
Kristen Bos (She/Her)
University of Toronto, Canada

Kristen Bos is the Co-Director of the Technoscience Research Unit and an assistant professor of Indigenous Science and Technology Studies at the University of Toronto. Kristen is an Indigenous feminist researcher trained in archaeological approaches to material culture as well as an Indigenous science and technology studies (STS) researcher, who is concerned the relationship between colonial, gendered, and environmental violence. Her current work examines and creates tools to navigate the relationships between data, pollution and colonialism in museum archives; Canada's Chemical Valley, where 40% of Canada's petrochemicals are processed; and in the accelerated creation of self-driving laboratories (SDLs), also called materials acceleration platforms (MAPs), using artificial intelligence (AI), robotics, materials sciences, and high-throughput chemistry. She is also the author of the upcoming novel, *The Interrogation Room* (Alchemy, 2025). Kristen is urban Métis based in Toronto, but her homeland is northern Alberta where prairie transitions into boreal forest.



Stephanie Dick (She/Her)
Simon Fraser University, Canada

Stephanie Dick is an Assistant Professor in the School of Communication at Simon Fraser University. She holds a PhD in History of Science from Harvard University. Stephanie is a historian of mathematics, computing, and artificial intelligence with an interest in the epistemological consequences of automation. In particular, her work explores the automation of mathematical intelligence, theorem-proving and problem solving as well as the introduction of computerized databases and algorithmic identification to law enforcement. She is co-editor with Janet Abbate of *Abstractions and Embodiments: New Histories of Computing and Society* (Johns Hopkins University Press); co-editor of *Histories of AI: A Genealogy of Power* (BJHS Themes); and she co-edits the "Mining the Past" Column at the *Harvard Data Science Review*. Before joining the faculty at SFU and back coming home to Canada, Stephanie was a Junior Fellow with the Harvard Society of Fellows and an Assistant Professor at the University of Pennsylvania.



Steve DiPaola (He/Him)
Simon Fraser University & RSC College, Canada

Steve DiPaola is a professor and AI researcher at Simon Fraser University. He came to SFU from Stanford University where he was a research faculty. Past Director of Cognitive Science at SFU, he is a pioneer in AI based computational creativity exploring the marriage of cognitive science and artificial intelligence-based modelling. His work in computational/cognitive modelling has had applications in science, e-health, education, and the arts. His expertise conjoins neuroscience, artificial intelligence with human mechanisms of creativity, expression, and empathy. He does this to both build: foundational cognitive theories on how humans are creative, how they express, have empathy; and to create new forms of computational, social, and artistic artifacts. His fine AI based artwork has been exhibited at major museums including the MOMA, Tate and Whitney museums. As a cognitive based AI scientist, he has over 100 scientific papers including his AI systems in empathetic 3D chatbot characters.

**Audrey Durand (She/Her)****Université Laval, Canada**

Audrey Durand is an Assistant Professor in the Department of Computer Science and Software Engineering at U. Laval. She holds a Canada CIFAR AI Chair, she is an associate academic member of Mila – Quebec AI Institute, and she is the leader of the Methods of Artificial Intelligence and Data Processing Axis of the Institute for Intelligence and Data at U. Laval. She is known for her theoretical and applied work in reinforcement learning and bandit algorithms, in addition to the applications of AI in general to health sciences. Her current research focus is on methods that impact their environment through their learning. She is well-involved in her community, having served as an Area Chair and having co-organized several workshops collocated with prestigious international conferences such as NeurIPS, ICML, and ICLR.

**Golnoosh Farnadi (She/Her)****McGill University and Mila Quebec AI Institute, Canada**

Golnoosh Farnadi is an assistant professor at the school of computer science at McGill University and a core academic member at Mila (Quebec AI Institute). The theme of her research is responsible AI and her recent work has mainly focused on addressing privacy and algorithmic discrimination in Machine learning and decision making models. Her research has garnered over 1000 citations, and she has won two paper awards. In 2021, Golnoosh was appointed a Canada AI CIFAR chair for her work on algorithmic fairness in AI. Golnoosh was one of the organizers of the 1st Mila/IVADO summer school on Bias and Discrimination in 2018 in Montreal and has been the scientific director of the online MOOC based on its content. She was an organizer of the NeurIPS workshop series on algorithmic discrimination from 2020-2023 and the NeurIPS 2022 tutorial on algorithmic fairness at the intersection and the senior program chair of the Montreal AI Symposium 2022. Golnoosh co-authored a whitepaper with UN-Habitat on AI's risks, applications, and governance in cities in 2022 and she published a book on missing links in AI governance, as part of a collaborative effort between Mila and UNESCO in 2023. Golnoosh is a recipient of the 2021 Google Research Scholar Award, the 2021 Facebook Research Award on privacy-preserving technology, was named one of the 2022 Rising Stars in AI Ethics, the 2022 Google Research Excellence Award, Finalist of the WAI 2023 responsible AI leader of the year award, the 2023 Google Award for Inclusion Research, and was named one of the 100 Brilliant Women in AI Ethics in 2023.

**Tom Fleming (He/Him)****Arctoris, UK**

Tom Fleming MChem FRSA FRSB FRSC, Co-Founder & COO, has a background rooted in organic chemistry with experience spanning early-stage drug discovery, chemical biology, cheminformatics, and laboratory automation. His journey started with a simple fascination with molecular structures, which led to a deep passion for making a difference in healthcare and life sciences. Being named one of the 'Top 100 Most Influential People in Healthcare and Lifesciences' was an honour, as it symbolised a recognition of his team's hard work and dedication. As a Fellow of the Royal Societies of Biology, Chemistry, and Arts, Manufactures, and Commerce, he feels proud to be part of communities striving to push the boundaries of knowledge and innovation. As an inaugural member of the UK Young Academy, he is determined to help make the UK a Science Superpower and maximise the societal impact of scientific progress.



Anthony Fuller (He/Him)
Carleton University, Canada

Anthony has a B.Eng in Aerospace Engineering, an M.A.Sc. in Systems and Computer Engineering, and is currently a PhD student at Carleton University. He is broadly interested in deep learning, particularly self-supervised learning, vision transformers, multimodal applications, and interpretability. Anthony enjoys learning about deep learning, his book club, and playing with his nieces and nephews.



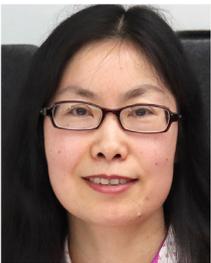
Alona Fyshe (She/Her)
University of Alberta, Canada

Alona Fyshe is an Associate Professor with a joint appointment in the Computing Science and Psychology Departments at the University of Alberta. She is fellow at the Alberta Machine Intelligence Institute (Amii) and holds a Canada CIFAR AI Chair. Alona received her BSc and MSc in Computing Science from the University of Alberta, and a PhD in Machine Learning from Carnegie Mellon University. Alona uses machine learning to analyze brain images collected while people read text or view images, which allows her to study how the human brain represents meaning. Alona also studies how computer models learn to represent meaning when trained on text or images. Alona leverages the connections between computer representations of meaning and those found in the human brain in order to advance our understanding of the brain, and the state of the art in machine learning.



Maria Galvez Trigo (She/Her)
Cardiff University, UK

Marisé is a Lecturer (Assistant Professor) in the School of Computer Science and Informatics at Cardiff University, and part of the Human-centred Computing Research Section and its Subgroup in Computational and Human-centred Robotics. She is interested in Assistive Robotics, Human-Robot Interaction, Human-Computer Interaction and applications of Machine Learning in those areas, with accessibility and co-design as key aspects of her research. Marisé has led and collaborated in research projects in those areas, and is currently Co-Investigator in the UK Robotics and Autonomous Systems Network Plus (UK-RAS+), leading initiatives for Early Career Researches and Equality Diversity and Inclusion.



Hui Guo (She/Her)
The University of Manchester, UK

Hui is a Reader in Biostatistics and Lead of the Centre for Biostatistics, University of Manchester. She studied for a PhD in Statistics and then worked as a research associate at Cambridge University. Her PhD research entitled “Statistical Causal Inference and Propensity Analysis” was within decision-theoretic framework. Her postdoctoral research was on the development and applications of statistical methodologies with the aim of understanding the genetics and mechanisms of common complex disease. Hui’s research focuses on statistical causal inference from observational studies. Her current work is on identification of causal genes of clinical outcomes from large-scale studies, optimising predictive models and investigating biological causal pathways and/or networks of diseases using multi-omics data.



Victor Gutierrez Basulto (He/Him)

Cardiff University, UK

Victor has been working in AI and its application to semantic data management for over ten years. His research interest is in the development of intelligent systems with capabilities to reason in the presence of structured and unstructured expert and commonsense knowledge. Victor has carried out foundational work on ontology-enriched information systems capturing uncertainty and dynamic aspects of knowledge, and on data quality. More recently, he has work at the intersection of knowledge representation and machine learning. He has investigated the knowledge representation capabilities of knowledge graph completion (KGC) methods and developed KGC approaches that use techniques inspired from the field of knowledge representation and reasoning.



Christoph Haase (He/Him)

University of Oxford, UK

Christoph Haase is an Associate Professor at the Department of Computer Science at the University of Oxford and a Fellow of St Catherine's College Oxford. He obtained his PhD from Oxford University and has held appointments at Microsoft Research Cambridge, UK, the École Normale Supérieure de Cachan, France, and University College London, UK. In 2019, Christoph was a recipient of a Starting Grant from the European Research Council to work on the mathematical foundations of logical theories of arithmetic. He serves as Associate Editor for the Journal of Computer and System Sciences and is Editor in Chief of the Automata, Logic and Semantics Section of the journal Discrete Mathematics and Theoretical Computer Science.



Eleanore Hickman (She/Her)

University of Bristol & UK Young Academy, UK

Eleanore is a lecturer at the University of Bristol where she teaches corporate governance, corporate law simulation and corporate finance. She is a director of Bristol's Centre for Private and Commercial Law and a member of the UK Young Academy. Eleanore researches the responsibilities of boards and senior management in companies and financial institutions, both organisationally and in relation to the societies they operate in. In this context she is interested in how the use and development of AI should be regulated. Eleanore has published in leading legal journals and her monograph entitled "Diversity, Merit and Power in the C-suite" was published in 2023. She holds a PhD, LLM and LLB in Law from University College London and has undertaken post-doctoral studies at the University of Cambridge. She is a non-practising solicitor, having trained and practised at Weil, Gotshal & Manges.



Karim Jerbi (He/Him)

UNIQUE Center & Université de Montréal & RSC College, Canada

Karim Jerbi is a professor at Université de Montréal. He holds the Canada Research Chair in Computational Neuroscience and Cognitive Neuroimaging and is the director of the UNIQUE Centre, the Quebec Neuro-AI research center. He is a member of the Royal Society of Canada's College of New Scholars, Artists and Scientists. He obtained a PhD in Cognitive Neuroscience and Brain Imaging from the Pierre & Marie Curie University in Paris (France) and a biomedical engineering degree from the University of Karlsruhe (Germany). His research lies at the crossroad between cognitive neuroscience and AI. Ongoing projects in his lab use electrophysiological brain recordings to examine large-scale brain network dynamics in a range of cognitive processes (e.g. decision-making) and across different states of consciousness (resting wakefulness, sleep, dreaming, anesthesia, meditation and psychedelic states). Dr Jerbi also has a keen interest in the convergence between AI, creativity and art.



Argyro Karanasiou (She/Her)
University of Birmingham, UK

Dr Argyro Karanasiou is Assistant Professor in Law and Innovation at the University of Birmingham, where she leads on the MSc in Responsible Data Science. Her work is contributing to the growing body of 'law and emerging technologies' transdisciplinary scholarship and has earned her visiting research affiliations with Yale Law School (ISP Alumna), New York University (ILI Alumna), Harvard Law (affiliate Faculty staff CopyX), and Complutense Madrid (ITC Erasmus Fellow). She is actively involved in several technology policy initiatives in the UK and worldwide, having served as a contracted consultant for the Council of Europe, and as an OSCE expert for Online Media She has further contributed invited expert insights on several occasions, most notably for the International Observatory on Information and Democracy (2023), the Equality and Human Rights Commission (2020), the Chatham House (2018), the US Air Force (2018), the Royal Society (2016), and the Electronic Frontiers Foundation (2016).



Atoosa Kasirzadeh (She/Her)
University of Edinburgh and Alan Turing Institute, UK

Atoosa is a philosopher, mathematician, and systems engineer. She is an assistant professor (Chancellor's Fellow) in the philosophy department and the Director of Research at the Centre for Technomoral Futures at the University of Edinburgh, and a Research Lead at the Alan Turing Institute. Prior to this, she held research positions at DeepMind and Australian National University. She has a PhD in philosophy of science and technology (2021) from the University of Toronto and a PhD in mathematics (2015) from the Ecole Polytechnique of Montreal. Her current research is focused on ethics, safety, and philosophy of AI (value alignment, interpretability, generative models, recommender systems).



Marlos C. Machado (He/Him)
University of Alberta, Canada

Marlos C. Machado is an Assistant Professor at the University of Alberta and a Fellow of the Alberta Machine Intelligence Institute. He is also a principal investigator of the Reinforcement Learning and Artificial Intelligence group and a Canada CIFAR Chair in Artificial Intelligence. Before joining the University of Alberta he was a research scientist at DeepMind and Google Brain for four years. Marlos' work has been published in top venues in machine learning and artificial intelligence, including Nature, JMLR, JAIR, NeurIPS, ICML, and ICLR. In particular, Marlos has introduced ideas such as stochasticity in the popular Arcade Learning Environment, temporally-extended exploration through options, and the application of deep reinforcement learning to control Loon's stratospheric balloons. His research has also been featured in popular media such as BBC, Bloomberg TV, and Wired.



Nicole Mwananshiku (She/Her)
The Royal Society, UK

Nicole Mwananshiku is a Policy Advisor in the Data and Digital Technologies team at the Royal Society. The team focuses on how data-driven technologies and artificial intelligence can, and should, be used to benefit humanity. She works on the following projects: Science in the Age of AI (exploring how AI is changing the nature and method of scientific research) and The online Information Environment (exploring how the internet and data-driven technologies affects the production of disinformation). She holds an masters degree in Conflict, Security and Development from the University of Exeter.



Isar Nejadgholi (She/Her)
National Research Council, Canada

Dr. Isar Nejadgholi is a Senior Research Scientist at the National Research Council Canada and an Adjunct Professor at the University of Ottawa. With an extensive background in deploying machine learning systems across diverse use cases in industry and academia, Isar brings an interdisciplinary approach to her research. She advocates for the responsible design and use of AI systems to drive positive societal changes. Her current work mostly focuses on potential applications of language technologies, specifically large language models, in disaster response, countering hateful language in online platforms, and newcomer settlement services.



Denis Newman-Griffis (They/Them)
University of Sheffield & UK Young Academy, UK

Denis Newman-Griffis is a Lecturer in Data Science in the University of Sheffield Information School and Co-Chair of the UK Young Academy. Their research investigates the principles and practices in developing responsible and ethical AI to advance human well-being, with a particular focus on health and disability. This work spans the development and evaluation of natural language processing technologies for healthcare, critical analysis of disability in AI systems, and building organisational best practice for responsible AI. Denis is a Research Fellow of the Research on Research Institute and PI of the GRAIL project studying responsible AI and machine learning in research funding (<https://researchonresearch.org/project/grail/>). They are an active educator in data science at undergraduate and postgraduate level and pursue pedagogical innovation in teaching contemporary data science and AI. Denis is an EDI leader in biomedical informatics and recipient of the American Medical Informatics Association's 2021 Doctoral Dissertation Award.



Rita Orji (She/Her)
Dalhousie University & RSC College, Canada

Prof. Rita Orji is a Canada Research Chair in Persuasive Technology and a Computer Science Professor at Dalhousie University, Canada where she directs the Persuasive Computing Lab. Her research at the intersection of technology and human behaviour focuses on designing interactive technologies to empower people, improve lives, and contribute to solving many societal problems. Specifically, technologies that integrate into people's daily lives and support them to achieve various self-improvement goals. She applies her work to tackle real-life problems in various domains including improving a wide range of health and wellness objectives such as mental health, healthy eating, physical activity, smoking cessation, sexual and other health risk behaviours. She is a recognized STEM diversity ambassador who is passionate about inspiring the next generation of youths and female tech leaders, promoting research excellence, equity, diversity, and inclusion in STEM. Prof. Orji also consults and serves as a board member for many agencies.



Alina Patelli (She/Her)
Aston University & UK Young Academy, UK

Dr Patelli holds a PhD in computer science awarded by Aston University in 2017, and one in systems engineering awarded by her Romanian alma mater in 2011. She specialises in evolutionary computation, a type of biologically-inspired Artificial Intelligence. Dr Patelli's research focus is on genetic programming with transfer learning and its applications in smart cities, specifically traffic modelling and prediction. Dr Patelli is also interested in autonomic, knowledge-based systems, and self-adaptation and self-organisation in computing.



Isabel Pedersen (She/Her)
Ontario Tech University, Canada

Isabel Pedersen, PhD, is Director of the Digital Life Institute and Professor of Communication and Digital Media Studies at Ontario Tech University. She held the Canada Research Chair in Digital Life, Media, and Culture (Tier 2) from 2012-2022. Dr. Pedersen has made numerous original SSHRC-funded research contributions on the social implications of autonomous AI agents, social robots, and generative AI. She is co-author of two recent books, *Writing Futures: Collaborative, Algorithmic, Autonomous* (Springer, 2021) and *Augmentation Technologies and Artificial Intelligence in Technical Communication: Designing Ethical Futures* (Routledge, 2023). She has advised Policy Horizons Canada, Global Affairs Canada, and Pew Research Center, amongst other organizations.



Siva Reddy (He/Him)
McGill University and Mila Quebec AI Institute, Canada

Siva Reddy is an Assistant Professor in the School of Computer Science and Linguistics at McGill University. He is also a Facebook CIFAR AI Chair, a core faculty member of Mila Quebec AI Institute and a research scientist at ServiceNow Research. His research focuses on representation learning for language that facilitates reasoning, conversational modeling and safety. He received the 2020 VentureBeat AI Innovation Award in NLP, and the best paper award at EMNLP 2021. Before McGill, he was a postdoctoral researcher at Stanford University and a Google PhD fellow at the University of Edinburgh.



Tulika Saha (She/Her)
University of Liverpool, UK

Dr Tulika Saha is a Lecturer (Assistant Professor) at the Department of Computer Science in the University of Liverpool (UoL), United Kingdom. She is a part of the Natural Language Processing group at UoL. Prior to joining UoL, she worked with Professor Sophia Ananiadou, as a postdoctoral researcher at the National Centre for Text Mining (NaCTeM), University of Manchester, UK. She earned her PhD in 2021 from Indian Institute of Technology Patna, India. Her PhD supervisors were Dr Sriparna Saha and Professor Pushpak Bhattacharyya. Her research focuses on developing end to end dialogue systems in multi-lingual and multi-modal framework. Independent target is set on addressing all the primary modules of Dialogue System which includes Natural Language Understanding, Dialogue Management Strategy and Natural Language Generation using Deep Learning and Reinforcement Learning. Several aspects of human behavior such as emotion and sentiment are incorporated to make the Virtual Assistant efficient. Investigations have also been extended to analyse the content on social media platform such as Twitter. She has recently also forayed her investigations in NLP for Social Good such as healthcare, education, law and climate etc.



Ervin Sejdić (He/Him)

University of Toronto and North York & RSC College, Canada

Dr. Ervin Sejdić received B.E.Sc. and Ph.D. degrees in electrical engineering from the University of Western Ontario, London, Ontario, Canada in 2002 and 2008, respectively. From 2008 to 2010, he was a postdoctoral fellow at the University of Toronto with a cross-appointment at Bloorview Kids Rehab, Canada's largest children's rehabilitation teaching hospital. From 2010 until 2011, he was a research fellow at Harvard Medical School with a cross-appointment at Beth Israel Deaconess Medical Center. From 2011 until 2021, he was a faculty member at the University of Pittsburgh. In 2021, he joined the University of Toronto as a faculty member. He is also the Research Chair in Artificial Intelligence for Health Outcomes at North York General Hospital in Toronto. From his earliest exposure to research, he has been eager to contribute to the advancement of scientific knowledge through carefully executed experiments and ground-breaking published work. This has resulted in co-authoring over 150 publications. In February 2016, President Obama named Dr. Sejdić as a recipient of the Presidential Early Career Award for Scientists and Engineers. In 2017, Dr. Sejdić was awarded the National Science Foundation CAREER Award. In 2018, he was awarded the Chancellor's Distinguished Research Award at the University of Pittsburgh. Dr. Sejdić is the editor-in-chief of Biomedical Engineering Online, an associate editor for IEEE Transactions on Biomedical Engineering and Digital Signal Processing. Dr. Sejdić's passion for discovery and innovation drives his constant endeavors to connect advances in engineering to society's most challenging problems. Hence, his research interests include biomedical signal processing, gait analysis, swallowing difficulties, advanced information systems in medicine, rehabilitation engineering, assistive technologies and anticipatory medical devices.



Luke Stark (He/Him)

Western University, Canada

Luke Stark is an Assistant Professor in the Faculty of Information and Media Studies at Western University in London, ON and the Inaugural Scholar-in-Residence at the University of Toronto's Schwartz Reisman Institute for Technology and Society. He has previously been a Postdoctoral Researcher in the Fairness, Accountability, Transparency, and Ethics (FATE) Group at Microsoft Research, and a Fellow and Affiliate at the Berkman Klein Center for Internet & Society at Harvard University. Stark's scholarship on the ethical and social impacts of artificial intelligence systems has appeared in the proceedings of ACM conferences including ACM CHI, ACM FAccT and AAAI/ACM AIES, as well as in numerous academic journals. His current book project Reordering Emotion: Histories of Computing and Human Feeling from Cybernetics to Artificial Intelligence is under contract with the MIT Press and is the first comprehensive history of the topic. His work is funded by the Social Sciences and Humanities Research Council of Canada (SSHRC), and his scholarship has been cited by in publications including CNN, The New York Times, The Washington Post, The Guardian, The Atlantic, Maclean's, and CBC Radio/Television.



Ahmad Taha (He/Him)

University of Glasgow & UK Young Academy, UK

Dr Taha is a Lecturer in Autonomous Systems and Connectivity at the James Watt School of Engineering, University of Glasgow, with a decade of experience across esteemed higher education institutions spanning Egypt, the UK, and China. His research journey centres on cyber-physical systems for energy management and digital healthcare. Dr Taha has authored/co-authored over 35 publications showcased in reputable venues and has contributed as a Principal Investigator/Co-Investigator to research grants exceeding £350k in value. Dr Taha has earned recognition from the Royal Academy of Engineering through the Global Talent scheme, he is an inaugural member of the UK Young Academy and an academic advisor to the Commonwealth Scholarship Commission. Dr Taha holds the title of Fellow of Advanced Higher Education and is among the pioneering UK-CGE recognised Associate Supervisors. His journey began with a dual BSc with Honours from MSA University and the University of Greenwich. Securing scholarships for his MSc and PhD pursuits, he earned distinction in Embedded Systems in 2014 and completed a collaborative PhD in 2020 with Medway NHS Foundation Trust in Kent, UK.



Efi Tsamoura (She/Her)
Samsung AI Centre, UK

Efi Tsamoura is a Senior Researcher at Samsung AI, Cambridge, UK. In 2016, she was awarded an early career fellowship from the Alan Turing Institute, UK, and before that, she was a Postdoctoral Researcher in the Department of Computer Science of the University of Oxford. Her main research interests lie in the areas of logic, knowledge representation and reasoning, and neuro-symbolic integration. Her research has been published in top-tier AI and database venues (SIGMOD, VLDB, PODS, AAI, ICML, NeurIPS, etc.).



William J Turkel (He/Him)
University of Western Ontario & RSC College, Canada

William J Turkel is Professor of History at the University of Western Ontario and internationally recognized for his innovative work in digital history. He uses machine learning, text mining, and computational techniques in his study of the histories of science, technology, and environment, drawing on many decades of programming experience. Author of *Spark from the Deep* (Johns Hopkins, 2013), *The Archive of Place* (UBC, 2007) and the open access textbook *Digital Research Methods with Mathematica*, (2nd ed 2019), he was Project Director of Digital Infrastructure for the SSHRC-funded Network in Canadian History & Environment (NiCHE) from 2004-14. Dr. Turkel is a member of the College of New Scholars, Artists and Scientists of the Royal Society of Canada (2018-25) and a recipient of the Wolfram Innovator Award (2020).



Karina Vold (She/Her)
University of Toronto, Canada

Dr. Karina Vold is an Assistant Professor at the Institute for the History and Philosophy of Science and Technology at the University of Toronto. She is also a Research Lead at the U of T Schwartz Reisman Institute for Technology and Society, an AI2050 Early Career Fellow with the Schmidt Futures Foundation, a Faculty Associate at the U of T Centre for Ethics, and an Associate Fellow at the University of Cambridge's Leverhulme Centre for the Future of Intelligence. Vold specializes in Philosophy of Cognitive Science and Philosophy of Artificial Intelligence (AI), and her recent research has focused on human autonomy, cognitive enhancement, extended cognition, and the risks and ethics of AI.



Ivan Vulić (He/Him)
University of Cambridge, UK

Ivan Vulić is a Principal Research Associate and a Royal Society University Research Fellow in the Language Technology Lab, University of Cambridge. He is a member of the Steering Committee of the newly established Centre for Human Inspired Artificial Intelligence at Cambridge. Ivan holds a PhD in Computer Science from KU Leuven awarded *summa cum laude*. In 2021 he was awarded the annual Karen Spärck Jones Award from the British Computing Society for his research contributions to Natural Language Processing and Information Retrieval. His core expertise is in representation learning, applied machine learning, conversational AI, multilingual and multi-modal NLP and human language understanding. He has published numerous papers at top-tier NLP and IR conferences and journals, and his research work also resulted in several best paper awards. He serves as an area chair and regularly reviews for all major NLP and Machine Learning conferences and journals. Ivan has given numerous invited talks at academia and industry, and co-organised a number of NLP conferences



Erin Young (She/Her)

The Alan Turing Institute, UK

Dr Erin Young is a Research Fellow and Project Co-Lead in the Public Policy Programme at The Alan Turing Institute, the UK's National Institute for Data Science and AI. She has a PhD (DPhil) from the University of Oxford, where she studied socio-technical practices of interdisciplinary R&D projects. Previously, Erin held positions at the United Nations in Paris, Kantar/WPP in London, Thomson Reuters in New York City, and Stanford University. She holds an MSc (Distinction) from the University of Oxford in Education (Learning and Technology), a PGC in International Business Practice, Finance and Organisational Behaviour, and a BA from the University of Cambridge. Erin is passionate about tackling pressing societal and economic issues and generating responsible impact with AI by forging strong relationships internationally between government, academia and the private sector. She regularly lectures, advises and collaborates across institutions including European Parliament, GCHQ, the Cabinet Office, Bank of England and British Embassies/Consulates globally.

Partners and Leadership

Elissa Strome (She/Her)

Executive Director, Pan-Canadian AI Strategy, CIFAR

Michel Tremblay (He/Him)

President, Academy of Science
Royal Society of Canada

Organizing Committee

Ashleigh Carver (She/Her)

Senior Scientific Programmes Officer,
International
The Royal Society

Amelia Domaradzki (She/Her)

Programmes Officer,
The Royal Society of Canada

Luke Clarke (He/Him)

Head of International Affairs,
Americas/Multilaterals/Africa,
The Royal Society

Russel MacDonald (He/Him)

Programmes Manager,
The Royal Society of Canada

Grandma Karen MacInnis (She/Her)

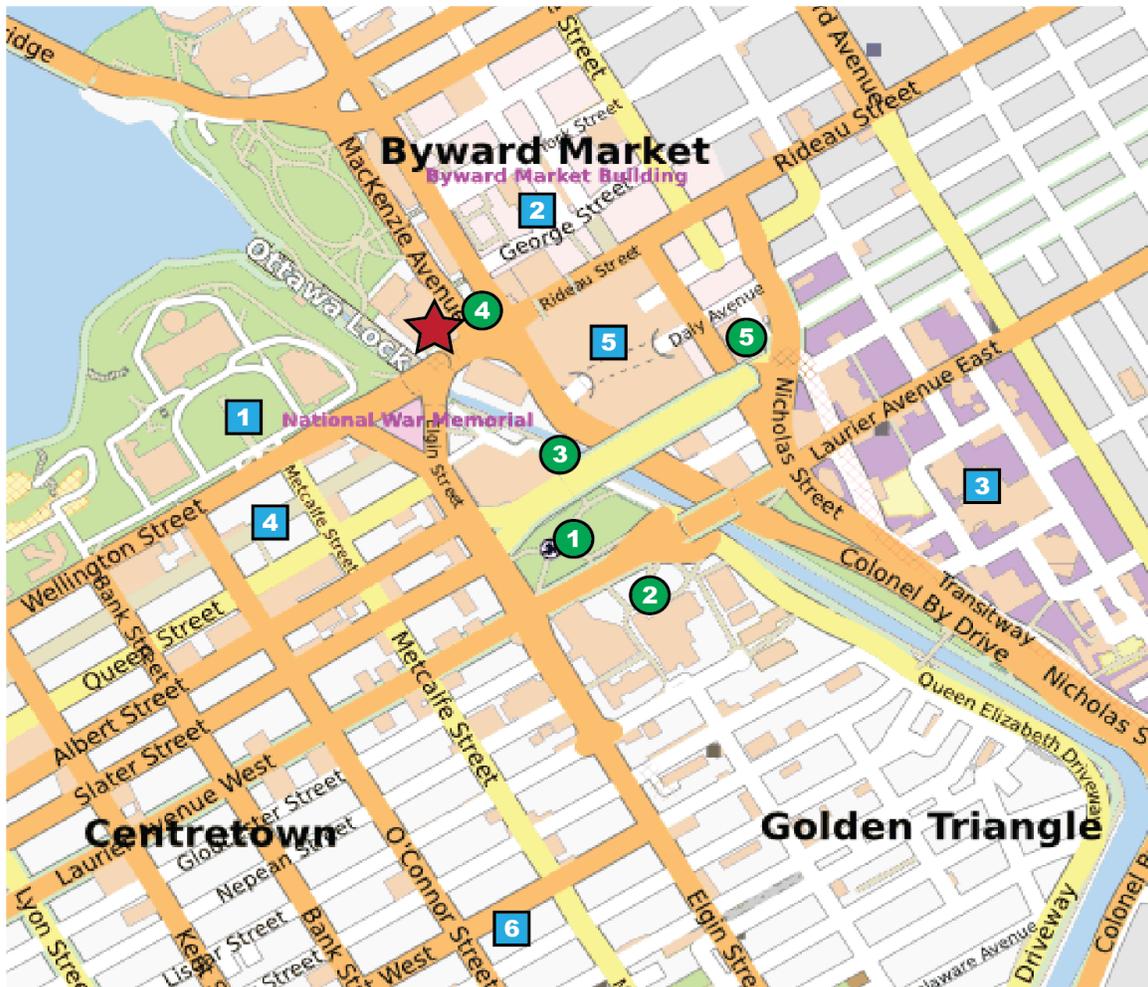
Knowledge Keeper and Elder

Karen, (Nenookaasiwag kwi), I am Ojibwa from Walpole Island First Nation (Bkejwanong).

“While living in the Timmins area, I spent many years working across the North as a therapist and traditional healer both on and off reserve using Indigenous healing & western psychotherapy.

In 1998 I moved to the unseeded territory of the Algonquin people (Ottawa) as a therapist, healer, knowledge keeper and Elder, I work with both Aboriginal and non-Aboriginal organizations, providing guidance, around traditional roles, teachings and ceremonies for both individuals, groups and organizations.

Sites Map of Ottawa



- ★ **Chateau Laurier**
- ① **Confederation Park (Ice sculptures, Tues. eve)**
- ② **Rink of Dreams (Skating, Tues. eve)**
- ③ **Rideau Canal Access (Tues. eve)**
- ④ **Metropolitan Brasserie Restaurant (Tuesday dinner)**
- ⑤ **Ottawa Art Gallery (Reception, Wed. eve)**
- ① **Parliament Hill**
- ② **Byward Market**
- ③ **University of Ottawa**
- ④ **Sparks Street**
- ⑤ **Rideau Centre**
- ⑥ **Royal Society of Canada Headquarters**

